

Safeguards for donors

Clashing perspectives on the ethics of the donation of human eggs for research purposes are likely to complicate international collaboration — whether stem-cell researchers like it or not.

What price a human egg? The question provokes a variety of emotions and responses. Some will argue that an egg has no monetary value when it is just one of those ovulated each month by billions of women and that perishes unfertilized. Others might contend that the same egg is priceless — because it could, if introduced to the correct sperm, form the seed of a new person. Others still will find it morally problematic even to pose the question, on the grounds that it treats human cells as merchandise.

But the question is being asked, nonetheless — indeed, it is at the forefront of a vigorous debate about whether and how much to pay women for donating their eggs for research (see page 606). Some biologists are keen to use fresh human eggs for the production of human embryonic stem cells. Yet they are anxious to distance themselves from the path taken by disgraced South Korean researcher Woo Suk Hwang, who lied about paying women for eggs and tainted the whole idea of oocyte donation.

In a Commentary in this week's issue, Insoo Hyun of Case Western Reserve University in Cleveland, Ohio, who chairs a task force exploring the issue at the International Society for Stem Cell Research, argues that women who donate eggs should be financially compensated for their time, discomfort and trouble (see page 629). Hyun argues that institutional review boards and other oversight bodies can select a level of remuneration that is enough to ensure that donors are compensated for their trouble, but not enough to blind them to the risks involved.

That will be an exquisitely fine line to draw. A sum that will be merely compensation for one woman could be enticement for another, poorer one. And in the real world, where questions over clinical research, informed consent and conflicts of interest have lately enjoyed a high profile, confidence in the ability of these review boards to ensure adequate oversight will not be universally shared.

The situation is further complicated by the health risks that may be posed by ovarian stimulation (see page 607). There are hints, but no definitive evidence, that the drugs used to stimulate the ovaries for both *in vitro* fertilization (IVF) and egg donation increase the long-term risks of cancer. How, then, can a fair level of compensation be set for risks that are essentially unknown?

One possible way forward lies in a practice called egg-sharing, in which women who are already considering or undergoing IVF are asked to contribute surplus eggs to research in return for treatment at reduced cost. The North East England Stem Cell Institute in Newcastle Upon Tyne, UK, will soon start offering this scheme. Proponents point out that, by working with donors who are using ovulation-stimulating drugs anyway, this avoids exposing otherwise healthy women to any risks associated with them. The idea does not, however, impress those who object on ethical grounds to any financial inducements to donors.

The International Society for Stem Cell Research is currently drafting guidelines on this issue. But they won't overrule local laws and regulations, the uneven application of which looks set to slow down stem-cell research. Researchers in nations that prohibit payment for eggs, for example, may find themselves unable to work on stem cells derived from eggs secured by collaborators elsewhere from paid donors.

In April 2005, the US National Academies issued guidelines stating that women who donate eggs for research should receive only direct expenses, such as travel to the clinic. If stem-cell researchers are to suggest that payments should be more extensive, they will have to make a more convincing case that adequate safeguards will be in place to protect donors' rights. ■

"How can a fair level of compensation be set for risks that are virtually unknown?"

Capturing carbon

Sequestration of greenhouse gases could play an important role in capping emissions.

Fresh approaches to energy use and production are vital if serious climate change is to be averted and developing countries are to attain the standards of living to which they aspire. However, the rich nations spend a deplorably low proportion of their research funds on energy — far less, in real terms, than they were spending 25 years ago.

The case for greater emphasis on energy research is overwhelming, and was made again last week by Martin Rees, president of Britain's Royal Society (see *Science* **313**, 591; 2006). But sometimes research

can get in the way of deployment. Scientists can always find further interesting questions, and research can become an end in itself. In some fields, there is a need, instead, for action. Energy conservation is the most obvious case. Carbon capture and storage — which offers the possibility of using fossil fuels without releasing carbon dioxide into the atmosphere (see page 620) — is another.

Bringing carbon sequestration onto a faster track requires more than scattered demonstration projects and a vague hope that prudent industries might voluntarily adopt it at some point in the future. The International Energy Agency predicts that 1,400 gigawatts of new, coal-fired generating capacity will be commissioned worldwide in the next 25 years. The United States has proposals for 153 coal-fired plants under consideration, few of which are likely to be designed with carbon capture in mind. And every year, China builds coal-powered plants capable of generating a stunning 75 gigawatts



— an energy project on a scale unprecedented in human history. To ensure that carbon dioxide from at least some of these plants is stored away in geologically suitable repositories requires more than research; it needs political will.

Evidence of such political will would include regulations or fiscal incentives to design plants so that carbon-capture equipment can be retrofitted to them with relative ease. And those who build plants must be convinced that, at some time in the future, any carbon dioxide they emit will be a cost to their businesses.

Carbon capture and storage is no panacea. It substantially decreases the efficiency of all existing plant types. It also requires an enormous infrastructure — putting carbon dioxide back down into the ground requires pipes and pumping comparable to that needed to bring oil and gas up out of it. Some reservoirs may turn out to be flawed, leaking carbon back over decades or centuries. Even under the most optimistic assumptions, less than half of human-produced carbon dioxide emissions could possibly be captured and stored.

Even so, carbon sequestration is the only credible option that would allow the continued use of fossil energy without the threat of dangerously altering Earth's climate system. Speeding up its deployment must therefore become a priority on the global energy agenda.

Parallel development of several different approaches to carbon

sequestration will be needed. The more hands-on experience that's gained with carbon capture from different plant types, and with carbon storage in different kinds of underground reservoir, the easier it will be to convince governments and industry to begin carbon sequestration on a commercial scale.

But political negotiations, regulatory frameworks and further research and development need not await the results of existing pilots. The G8 nations, together with China, India, Mexico, Brazil and South Africa, should tell their energy industries in no uncertain terms that carbon production will cost them, and that sequestration is a partial solution available in the short term. In some situations, subsidies and other incentives may be justified.

As the largest and fastest-growing emitters, respectively, the United States and China need to take the lead on this. Not all the approaches to carbon capture will work out, and some money will doubtless be wasted. But the risks are small compared with the potential benefits of making some significant inroads into carbon dioxide emissions. ■

"Carbon sequestration is the only option that would allow the use of fossil energy without the threat of dangerously altering Earth's climate system."

Toronto crossroads

The international AIDS meeting still has a purpose.

It has become fashionable over the years to knock the International AIDS Conference, which opens this weekend in Toronto. The usual critique is that the meeting is too messy, too noisy and lacking in compelling, new science. But these complaints miss the point of a unique biennial gathering that brings scientists, activists and drug-company and government officials into close and often contentious contact.

Laurie Garrett, a former journalist and a fellow at the Council on Foreign Relations in New York, has summed up the case for the prosecution. "What began in 1985 as an annual gathering of scientists, aimed at sharing laboratory findings and information from the battlefronts in the war on HIV, has been transformed into a meeting of 17,000 consultants, bureaucrats and activists fighting one another for money to build a huge global AIDS treatment program, employing tens of thousands of people," she wrote in *The New York Times* during the last meeting, in Thailand in 2004.

It is certainly true that the AIDS meetings have changed over the years, and that would-be participants must, from time to time, question the value of their attendance.

But it is vitally important that science continues to be represented in this forum. The need for science to inform and lead the fight against global AIDS has never been greater. This goes far beyond the laboratory research required to overcome the obstacles that have stymied the search for an AIDS vaccine. It extends to the need to help steer the allocation of resources in fighting the AIDS epidemic, which took some 2.8 million lives last year.

To give one example, policy-makers and scientists were worried a

few years ago that the availability of drug treatments in sub-Saharan Africa would lead to a rise in drug resistance and increases in risky behaviours. Some even used this fear as a reason for dragging their feet over the delivery of the drugs to people in poor countries.

But as we report on page 617 of this issue, scientists working on treatment roll-out programmes have found that patients in Africa adhere to their regimens just as well as those elsewhere. The main impediment to effective drug treatment in sub-Saharan Africa is not drug resistance, it turns out, but rather interruptions to drug supplies caused by logistical and financial issues that governments and donor agencies need to address.

At the Toronto meeting, researchers will also present studies on other important questions, such as the effectiveness of abstinence education and of condom promotion in AIDS-prevention programmes. They will discuss the challenges of starting clinical trials to test whether drugs can protect exposed people from contracting HIV — an approach strongly opposed by some activists — and consider the difficulties in monitoring treatment effectiveness and drug resistance in poor countries.

Discussions such as these can have a much greater impact on policy if they occur at a forum such as the Toronto meeting, in front of audiences that contain activists and government officials, as well as researchers. As scientists in this field are well aware, it is almost impossible to fully disentangle activism and politics from science in AIDS. That is why the scientists' presence at the AIDS meeting is so important.

The biennial ritual of activists heckling drug-company officials or tearing down their stands may seem trite, the political speeches tiresome, and the appearances by Hollywood figures and other celebrities frivolous. But that's the world we live in. Full participation in the AIDS meeting will, as it has in the past, serve to invigorate researchers and ensure the continued relevance of their work. ■

RESEARCH HIGHLIGHTS

Spot the difference

Phys. Rev. E **74**, 011914 (2006)

Jungle cats have true spots only when they are kittens. As they grow, the spots become rosettes — broken rings in leopards (pictured) and polygons in jaguars. This changing pattern is the latest to be successfully described using Turing models.

Alan Turing suggested in 1952 that biological patterns could be generated by two chemicals diffusing between cells and interacting under the animal's coat. Sy-Sang Liaw of National Chung-Hsing University in Taichung, Taiwan, and his team adjusted parameters in Turing's reaction-diffusion equations to create spots. They then tweaked the parameters so that the patterns resembled the coats of middle-aged big cats. However, no one has found the chemicals, which Turing called morphogens, that might make this model work in mammals.



L. ARNDT/NATUREPL.COM

BIOCHEMISTRY

DNA redesign

J. Am. Chem. Soc. doi:10.1021/ja062548x (2006)

Few would dispute the genius of DNA's chemical design. But some do question why its backbone evolved to be made from chains of five- rather than six-membered rings, when the latter might more easily be derived from common sugars, such as glucose.

Since the idea was first raised in the early 1990s, chemists have suspected that sugars' hexose rings might simply be too bulky to fit into DNA's neat structure. At last, Martin Egli of Vanderbilt University in Nashville, Tennessee, and his colleagues have confirmed this experimentally.

They studied the crystal structure of double-stranded homo-DNA, which has hexose in the backbone in place of DNA's deoxyribose. The result was a "slowly writhing ribbon", the team reports, with irregular twists and steps between base pairs.

PLASMA PHYSICS

Particle vision

Phys. Rev. Lett. **97**, 045001 (2006)

Protons act as photographers in a study that could lead to better imaging of compressed plasmas, such as those found in nuclear weapons or in some kinds of proposed fusion power plants.

In principle, energetic proton beams can take ultra-fast snap shots of the plasma, making it possible to study the material's evolution over time. Andrew MacKinnon of the Lawrence Livermore National Laboratory, California, and his colleagues present the first images taken using the technique.

The team created their pictures by firing a powerful laser at a thin foil of tungsten. The laser sent protons flying through an imploding plastic capsule into a detector.

IMMUNOLOGY

A spoonful of sugar

Science **313**, 670–673 (2006)

Researchers in the United States have revealed a longstanding paradox in immunology to have sugar-coated roots.

The antibody immunoglobulin G (IgG; pictured green below) plays an important role in activating inflammation to fight invaders but, puzzlingly, it can also soothe inflammatory autoimmune diseases when injected into the bloodstream.

Jeffrey Ravetch of the Rockefeller University, New York, and his colleagues report that IgG switches between these roles by altering the sugars attached to one region of the molecule. IgG is anti-inflammatory when it carries side chains of the sugar sialic acid, but boosts inflammation when the sugar

residues are removed. The two forms might bind different receptors on immune cells.

Doctors may be able to artificially boost the fraction of sugary IgG in intravenous gamma globulin, used to treat autoimmune diseases.

GENETICS

On standby

Genome Res. doi:10.1101/gr.5147406 (2006)

The parasite that causes African sleeping sickness, *Trypanosoma brucei*, devotes more than three-quarters of some chromosomes to standby genes, which it recruits to help it evade the host's immune system. This finding, by researchers at the University of Cambridge, UK, stretches current ideas about how chromosomes are organized.

It also reinforces the notion that the search for a vaccine to provoke immunity against *T. brucei* is a fruitless task — the parasite has multiple copies of variant surface glycoprotein genes, which make the proteins that are recognized by the immune system.

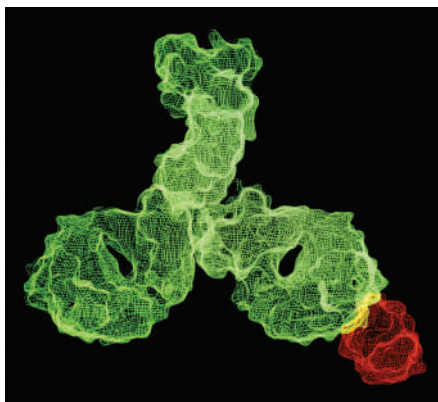
Sara Melville and her colleagues show that the regions in which the contingency copies are stored, just below the chromosomal tips, occupy huge tracts of the chromosome in some strains.

CELL BIOLOGY

Gene machine

Proc. Natl Acad. Sci. USA **103**, 12027–12032 (2006)

Hotspots that bind multiple regulatory molecules have been identified in the genome of the fruitfly *Drosophila*. These may act as storage sinks or assembly sites for 'machines' with more complex function, researchers in the United States and the Netherlands suggest.



J. C. REVY/SPL

The team studied a short portion of the genome, and noticed that transcription factors — which control the expression of sets of genes — congregated in distinct regions dotted along the section's length.

The binding hotspots may simply store the molecules. Another possibility is that the transcription factors interact to somehow draw together distant parts of the genome so that genes of similar function are close together. This would make their regulation easier to coordinate.

QUANTUM PHYSICS

Go with the grain

Science doi:10.1126/science.1130879 (2006)

When is a solid not a solid? When it's a supersolid — for frozen helium can flow like a superfluid, devoid of viscosity.

Researchers have suspected that the flow of supersolids — in which, like superfluids, the behaviour is dictated by quantum mechanics — is due to defects such as boundaries between helium-ice grains. Now Sébastien Balibar of the Ecole Normale Supérieure in Paris, France, and his colleagues present evidence for that.

They watched the level of solid helium-4 in a tube equilibrate with the reservoir in which the tube stood, indicating mass flow. Only solids full of grain boundaries can do this, suggesting that flow might occur in superfluid films at the grain surfaces.

IMMUNOLOGY

Flick the switch

Cell 126, 375–387 (2006)

The transcription factor NFAT may act as a sort of molecular switch, allowing an organism both to launch an immune attack on a foreign protein, and to suppress a response to a 'self' protein, suggest researchers.



NFAT is known to activate attacking lymphocytes when bound to its partner molecule, AP-1.

Now, Anjana Rao of Harvard Medical School in Boston, Massachusetts, Lin Chen from the University of Colorado, Boulder, and their colleagues have found evidence that NFAT also binds to another transcription factor, FOXP3. FOXP3 controls the function of specialized lymphocytes known as regulatory T cells, which prevent immune responses to the body's own cells and tissues.

PLANETARY SCIENCE

Snow on Mars

Astrobiology 6, 439–450 (2006)

The excitement surrounding recent reports of methane in the martian atmosphere centred on the gas's sources, which could conceivably be biological. The flip side to the puzzle is working out where the methane ends up. Understanding how the gas is degraded is important, both in estimating how much is formed and explaining the observed variation in its concentration.

Sushil Atreya of the University of Michigan, Ann Arbor, and his colleagues suggest that electrical discharges in storms and whirlwinds may play an important role. Such discharges could produce

large quantities of hydrogen peroxide, a powerful oxidant that will break down organic molecules. They calculate that such a mechanism could create 200 times the hydrogen peroxide produced by light-driven reactions, and may dust the surface of Mars (pictured above) with a peroxide 'snow'.

NEUROBIOLOGY

Brain sprouts

Nature Neurosci. doi:10.1038/nn1747 (2006)

Neuroscientists studying the way that brain cells make new connections face a chicken-and-egg conundrum. Many of the synapses in the cortex are junctions between cell branches known as dendritic spines. But which comes first, the spine or the synapse?

Graham Knott of the University of Lausanne in Switzerland and his colleagues argue that the spine sprouts first. They studied neurons in mice, imaging the appearance and disappearance of spines over 28 days, then using electron microscopy to examine tissue slices in more detail. Of the 57 spines identified, some of the youngest — seen only in the last day of *in vivo* imaging — lacked synapses.

This suggests that the spines grow 'naked', rather than swelling up under synapses that have already formed.

NASA/JPL/TEXAS A&M/CORNELL

JOURNAL CLUB

Rasmita Raval
University of Liverpool, UK

A surface scientist observes how self-seeking molecules build up asymmetry.

Nobel laureate Jean-Marie Lehn described supramolecular chemistry as 'molecular sociology'. Lehn, who won the 1987 chemistry prize for his work in this field, thus neatly encapsulated the concept of molecules congregating under the influence of multifarious

intermolecular forces.

One guiding force, important in biology, is the property of chirality, or 'handedness'. Chiral molecules exist in two mirror-image forms that cannot be superimposed. Some of us try to capture in our work the ease with which biological systems can distinguish chiral molecules or create chirality in reactions.

As a first step, we study how chirality can propagate from molecules to larger, supramolecular structures. Wolf Dieter-Schneider and co-workers at the Swiss Federal

Institute of Technology in Lausanne recently revealed a system of impressive complexity (M.-C. Blüm *et al.* *Angew. Chem. Int. Edn* 44, 5334–5337; 2005).

They used a scanning tunnelling microscope to image the structures formed by rubrene, a chiral molecule, on a gold surface. First it groups into pentagonal rings, resembling complex gearwheels. These wheels then link up into chains or form ten-membered rings. At each stage, intermolecular forces ensure that only molecules of the same

chirality assemble together.

It is striking that an essentially simple molecule — rubrene (C₂₄H₂₀) is a small, buckled sheet of carbon rings — can create such intricate homochiral architectures spontaneously.

The separation of molecules by chirality is a process that must have emerged at our very beginning, as proteins assembled from chiral amino acids in the primordial soup. Thus we, as molecular sociologists, are mapping the very first steps in the evolution of complex matter.

SPECIAL REPORT

Ethicists and biologists ponder the price of eggs

A shortage of human eggs hinders stem-cell research. Paying women to donate would increase supply, but experts are divided over the merits — and the long-term health consequences — of such a policy.

Should women be paid for the time, discomfort and health risks involved in donating eggs for research? The world's largest group of stem-cell scientists is grappling with the question, and has now asked the public for its views.

Stem-cell researchers want eggs so they can work on somatic cell nuclear transfer, or 'therapeutic cloning'. They hope to derive embryonic stem cells matched to patients' DNA, by transferring the nucleus of one of the patient's cells into a human egg and developing it into an embryo from which cells can be derived. The technique has great medical potential — but researchers are far from achieving it, and the main limiting factor in the research is the availability of human eggs to practise on.

So far, scientists have relied on women already undergoing fertility treatment donating their extra eggs for research. But the supply is meagre. To help persuade them, several labs are increasingly offering financial rewards, such as cheaper fertility treatment. Others are starting to ask healthy women to donate — triggering a debate about how such women should be compensated.

Some ethicists argue that women should receive compensation for the discomfort and effort involved. Others are worried that this will create an undue incentive that will coerce women — especially poorer ones — into giving up their eggs. The fact that so little is known about the long-term health risks of the procedure further complicates the picture (see 'Health effects of egg donation may take decades to emerge', opposite).

Scientists in many parts of Europe, Asia, the Middle East and North America ask women to donate eggs for research, but they all treat the practice differently. Scientists at the North East England Stem Cell Institute announced on 27 July that they had got permission from the UK Human Fertilisation and Embryology Authority to pay part of the cost of *in vitro* fertilization treatments for women who donate eggs for research (see *Nature* 442, 498; 2006).

But other European countries that accept egg donation for research, such as Sweden, prohibit payment for anything other than direct expenses. Japan bans egg donation altogether

because of the risk of complications, whereas China, along with several other countries, does not specifically address the question.

In the United States, a National Academies panel recommended last year that women should be reimbursed only for direct expenses. This approach has been adopted by the California Institute for Regenerative Medicine, and by the state of Massachusetts, home of the Harvard Stem Cell Institute (see *Nature* doi:10.1038/news060605-6; 2006). Both institutes are asking healthy women to donate eggs for research. But even their policies differ — for instance, California law would allow paying the travel expenses of women who come from out-of-state to donate eggs, whereas it's not clear how Harvard will handle this question. "At the moment we're confining the search to this area," says spokesman B. D. Colen.

Scientists want international guidelines so that they can share materials without worrying about how they were derived. So a task force of the International Society for Stem Cell Research (ISSCR) is considering the compen-

sation question as part of a larger effort to draw up guidelines. The task force, made up of scientists, ethicists and lawyers from 14 countries, was convened last year after the revelation of scientific misconduct by South Korean stem-cell researcher Woo Suk Hwang: as well as faking evidence of 11 human embryonic stem-cell lines derived by therapeutic cloning, he lied about paying hundreds of women for eggs, and

H. MORGAN/SPL



Research volunteers or organ donors?

Much of the disagreement on compensation has arisen from the question of how egg donors should be viewed. Laurie Zoloth, a bioethicist at Northwestern University in Evanston, Illinois, who was on the ISSCR task force, argues that egg donation is similar to organ donation, and should therefore be free of financial considerations.

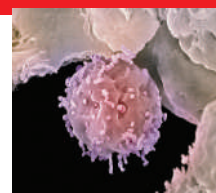
"One is asking the donor to undergo hyperstimulation, anaesthesia and minor surgery, then in a very real sense participate in the act of creation of an embryo that will be destroyed for research," she says. "This is a serious moral gesture, and I think it ought to be directed by a

serious and reflective moral decision."

The eggs are to be used for research rather than medicine — but, Zoloth argues, so were the first organ donations. If scientists start paying for eggs now, she says, it will be difficult to stop payments once the technique moves to the clinic.

But bioethicist Insoo Hyun of Case Western Reserve University in Cleveland, Ohio, who was also on the task force, argues that egg donors should be treated as research volunteers, like those who donate bone marrow for research (see Commentary, page 629). Such participants are generally reimbursed for their time and discomfort.

Meanwhile, Kevin Eggan of the Harvard Stem Cell Institute, who was not a member of the task force, has another concern. Although avoiding compensation may reduce the risk that women could be coerced into donating eggs, it would also restrict the practice to wealthy donors, who can afford child care or to take time off work. The eggs upon which research is carried out would therefore represent only the most privileged groups, and not the population as a whole. "I think it's very hard to hold the line that reimbursing for minimal expenses is the right thing to do for the women involved," he says. **E.C.**



DISCUSS

The pros and cons of egg donation on our Newsblog
<http://blogs.nature.com/news>

S. GSCHMEISSNER/SPL

Health effects of egg donation may take decades to emerge

In 1989, a healthy 32-year-old woman offered her infertile younger sister some of her healthy eggs, and with them the chance to have a baby. Doctors at the Cromwell IVF and Fertility Centre in London gave the donor hormones that made a batch of eggs in her ovary mature, and collected six eggs for fertilization. Three embryos were transferred to the younger sister and two were frozen. One baby girl was born.

Five years later, the doctors contacted the egg donor to ask whether to discard her frozen embryos. They discovered that she had been diagnosed with late-stage colon cancer that spread to her skull. She died just before her thirty-ninth birthday.

Doctors don't know if the fertility drugs caused or accelerated the woman's cancer. But the possibility prompted Cromwell infertility specialist Kamal Ahuja to report the case as a reminder of how little is known about the risks of donating eggs (K. E. Ahuja and E. G. Simons, *Hum. Reprod.* **13**, 227–231; 1998). "It shook us all up," he says.

Specialists in reproductive medicine say there is insufficient information about the long-term risks of drugs used to stimulate ovulation, a practice that has become more common in the past 25 years, with the proliferation of *in vitro* fertilization (IVF) and assisted reproduction. But some studies have suggested the drugs may be linked to the development of certain cancers.

The question is receiving renewed scrutiny now that scientists are asking healthy women to donate their eggs for stem-cell research — exposing them to the potential risks of ovulation stimulation without the end result of a baby (see Editorial, page 601). To discuss the issue, the California Institute of Regenerative Medicine (CIRM) in San Francisco has convened a meeting of experts to be held next month. Britain's Human Fertilisation and Embryology Authority (HFEA) will also tackle the issue in a forthcoming consultation on egg donation for research.

The uncertainty makes it even more difficult to reach a consensus on whether women who donate eggs should be compensated, and if so by how much (see 'Ethicists and biologists ponder the price of eggs'). "This discussion should emphasize long-term risk assessment rather than money," Ahuja says.

During ovulation stimulation for IVF or egg donation, women are given drugs that encourage the ovary to ripen several eggs simultaneously, rather than the one egg normally ovulated each month (see 'What egg donation involves', overleaf). Doctors know that this can have side effects ranging from moodiness to infection. The most serious is ovarian hyperstimulation syndrome, which seriously affects about 6% of women receiving the drugs. Thirty or more eggs start to develop at once and fluid leaks out of blood vessels and collects in the abdomen, causing nausea, bloating and very occasionally kidney failure or even death.

There is little information on how frequently ovulation stimulation has tragic side effects, says obstetrics and gynaecology professor Didi Braat of Radboud University Medical Centre in Nijmegen, the Netherlands, because doctors are often reluctant to report such cases and rarely have to. But deaths are thought exceptional: in a study reported at this year's meeting of the European Society for Human Reproduction and Embryology, Braat and her colleagues found only six deaths clearly linked to IVF from the medical records of some 100,000 women who underwent the procedure between 1984 and 2006.

So some specialists are more worried about the long-term risks of fertility drugs. In the

"This discussion should emphasize long-term risk assessment rather than money."

1990s, for example, studies pointed to a link between fertility drugs and breast or ovarian cancer, although it's not clear how cancer would be promoted. One study suggested that women who took an ovulation-stimulating drug called clomiphene citrate for more than a year had 11 times the risk of developing ovarian tumours compared with the general population (M. A. Rossing *et al.* *N. Engl. J. Med.* **331**, 771–776; 1994).

But these studies are controversial. It might be infertility, not fertility drugs, that predisposes women to disease. Other aspects of women's reproductive lives influence ovarian and breast cancer — pregnancy is thought to protect against tumours, for example. And ovarian cancer is so rare that it's hard to get a large enough sample to spot any connection.

Louise Brinton at the US National Cancer Institute in Bethesda, Maryland, and her colleagues tried to control for these factors in one

Mixed blessing? Donating eggs is a time-consuming and uncomfortable business.

obtained eggs from his female subordinates.

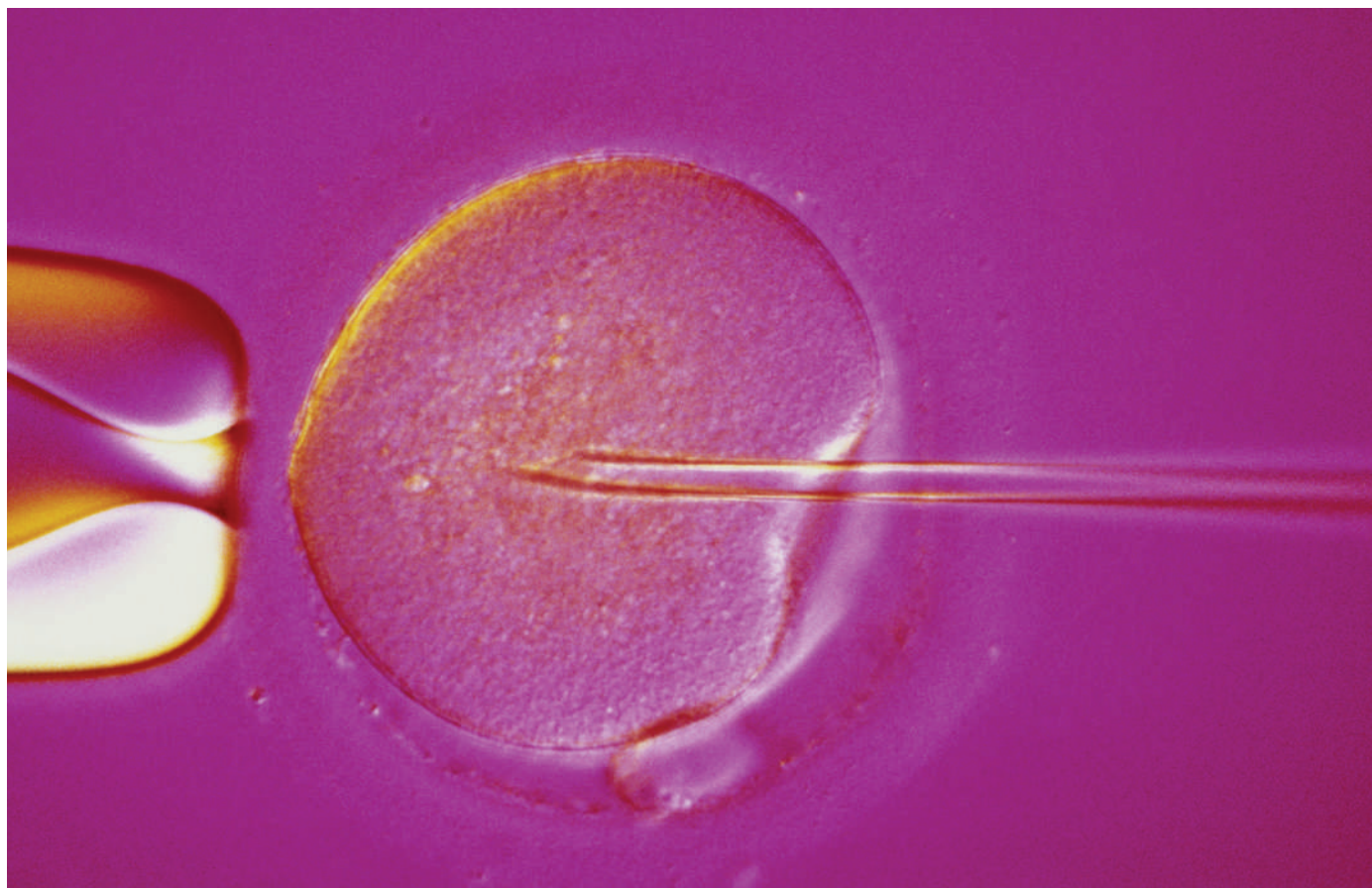
On 30 June, the ISSCR task force released draft guidelines at its annual meeting in Toronto. The guidelines embrace most of the principles proposed by the National Academies last year. But they differ on the issue of egg donation. The task force leaves the door open for a more liberal policy on compensation by stating simply that stem-cell research projects should be reviewed by a local oversight body, which must ensure "there are no undue inducements or other undue influences for the provision of human materials". What constitutes 'undue' is left to the local oversight bodies.

George Daley, a biologist at Harvard Medical School who chaired the task force, says that this is the best consensus the task force was able to achieve, because scientists and ethicists on the task force disagree so sharply about how egg donors should be treated (see 'Research volunteers or organ donors?').

The guidelines are seen as an important first step nonetheless, and are now open to public comment until 1 September, when the ISSCR will finalize the document.

"These are going to be seen as the rules set by scientists themselves, from the inside out," says Kevin Eggan of the Harvard Stem Cell Institute. "It's very useful for scientists to show that they have thought about these issues." ■

Erika Check



S. WALKER/GETTY IMAGES

Next generation: the techniques of assisted reproduction have become increasingly commonplace in the past 25 years.

of the most comprehensive studies so far. They collected the medical records of more than 12,000 women who received ovulation-stimulating drugs between 1965 and 1988. The team did not find statistically significant increases in breast and ovarian cancer, but did find that the women were around 1.8 times more likely to develop uterine cancer (M. D. Althuis *et al.* *Am. J. Epidemiol.* **161**, 607–615; 2005).

Such studies have reassured many specialists that the risks of ovulation stimulation are insignificant. But Brinton and others studying the issue say the picture is still incomplete. Brinton's study involved mainly women who took clomiphene citrate, rather than the gonadotropin hormones introduced for IVF in the 1980s. Researchers have only had a decade or so to study significant numbers of women taking the newer drugs, but extra cancers may not appear until the women reach 50 or 60.

Researchers also don't know whether studies on infertile women can be generalized to egg donors, who are typically younger and healthier. Epidemiologist Mary Croughan at the University of California, San Francisco, has unpublished data suggesting donors are at lower risk of cancer. But "it's important to fol-

low these women into the future", she says.

She and other experts want more extensive studies to follow up women who have had IVF treatment or donated eggs. There is at least one large study of the long-term effects of ovarian stimulation under way in the Netherlands. But it's unclear who will drive the effort, particularly when private fertility clinics may have little interest in finding out the potential risks of the drugs they use. Ahuja suggests that an

authority such as Britain's HFEA could coordinate such an investigation.

Some argue that the researchers asking women for eggs should help pin down the health risks. "There is some kind of ethical obligation there," says Mildred Cho of the Stanford University Center for Biomedical Ethics in California. The September CIRM meeting will focus on these risks and discuss what relevant data might be collected in the course of the institute's research.

"It's important for people to understand in the consent process that we don't know as much as we should about what those risks are," says Cho. Kevin Eggan of the Harvard Stem Cell Institute, for example, says that his group tells egg donors of the risks, and that they cancel the procedure if women show signs of ovarian hyperstimulation syndrome.

If the risks aren't made clear upfront, one well-publicized tragedy could kill efforts to find donors, adds John Buster, professor of obstetrics and gynaecology at Baylor College of Medicine in Houston, Texas. "If a woman has a cardiac arrest while giving eggs for stem-cell research, it won't go down too well."

Helen Pearson

What egg donation involves

- In one typical protocol, women take a gonadotropin-releasing hormone agonist daily for 1–2 weeks, which stops the pituitary from stimulating ovulation.
- They then inject gonadotropins such as follicle stimulating hormone to trigger the development of several egg-containing follicles.
- A third hormone triggers final maturation of the eggs.
- Eggs are collected with a needle inserted through the wall of the vagina into the ovary.



STRESSED-OUT PLANTS WARN THEIR OFFSPRING

Increased tendency to mutate is handed down to next generation.

www.nature.com/news

J. BURGESS/SPL

Oil from bombed plant left to spill

The environment has been caught in the cross-fire between Israel and Lebanon. So far, 15,000 tonnes of fuel oil has leaked into the Mediterranean Sea from Lebanon's bombed Jiyeh power plant, 30 kilometres south of Beirut.

With more than 100 kilometres of the Lebanese coastline, and 10 to 20 kilometres of Syria's coast now affected, biologists are scrambling to try to assess how much damage the Lebanese spill could do, and how best to contain it.

"Every oil spill is different," says Robin Crump, a marine biologist who was involved in the clean-up operations after a 1996 spill off the coast of Wales. "Everything depends on the type of oil, the temperature of the water and the air, the intertidal regions."

Satellite images show that most of the fuel oil leaked by the Jiyeh plant has been pushed to the Lebanese shore by southwesterly winds; the lack of wave action in the Mediterranean has meant the slick has not been broken up.

Although thick oil such as Jiyeh's is less immediately toxic to wildlife, it can cause problems in the long term. It can pick up sand and sediment from beaches and the mix can then be carried offshore and sink, spreading the oil over a wider area. Warm temperatures can also bake it into a biologically inert crust, like tar-mac, that protects the fresher oil underneath from breakdown by bacteria.

Although the Lebanese government has received international aid in the form of floating barriers and barrels of detergent to help combat the slick, clean-up has not yet begun. Part of the problem is that no emergency experts from outside the country have been granted security



A satellite image shows the extent of an oil slick, spilled from a bombed power plant in Lebanon.

clearance to conduct an impact assessment. And even if experts are given access, the conflict will make it hard to recruit locals to help clean the beaches, observers say.

An untreated spill can cause lasting harm, says Jacqueline Michel, president of Research Planning, an environmental-management firm based in Columbia, South Carolina. During

the first Gulf War in the early 1990s, oilfields were sabotaged and some 1.5 million tonnes of crude oil spilled into the Persian Gulf. Most of it ended up on the Saudi Arabian coastline, and only around 10% was ever recovered. Today, Michel says, "the intertidal communities have shown little or no recovery".

In Lebanon, the failure to mount a swift clean-up operation could most harm the region's threatened leatherback, loggerhead and green turtles. Bluefin tuna may also be at risk depending on where the oil goes, adds Luisa Colasimone of the United Nations Environment Programme. The eastern Mediterranean is home to a large spawning ground for the tuna.

Environmental damage can occur even after a hasty clean-up, says Charles Peterson, a marine ecologist at the University of North Carolina, Chapel Hill. He and his colleagues have studied the long-term effects of an oil slick spilled when the Exxon Valdez tanker ran aground in Alaska's Prince William Sound in 1989. In 2003 they reported that the recovery of whale, otter and duck populations at the site of the spill has been slower than expected owing to indirect effects of oil contamination (C. H. Peterson *et al. Science* **302**, 2082–2086; 2003). And a spill along the eastern coast of Panama caused extensive damage to the region's mangroves, despite efforts to minimize the impact of the slick.

On the other hand, some areas affected by spills bounce back relatively quickly. The 1996 spill off Wales is a case in point. "Ten years on, the system has recovered," Crump says. ■

Kerri Smith

With additional reporting by Michael Hopkin

DLR

Plan for postdoc union sparks backlash

Postdocs at the University of California, one of the most influential US university systems, are embroiled in a debate over whether they should unionize. Some back the action, saying it would give them much-needed leverage for improving their working conditions. But others are concerned about the way in which the union in question — United Auto Workers (UAW) — has recruited people to the cause.

If formed, the group would be only the second postdoc union in

the country, and by far the largest. The University of Connecticut Health Center in Farmington has had one since 2004, and it currently represents 125 postdocs. At the University of California, the UAW would act for some 5,800 postdocs.

UAW representatives, including graduate students and postdocs at the university, have gone from lab to lab collecting signatures in support of unionizing. The organization filed a petition to represent postdocs on 13 July. This month, some postdocs began a letter-writing campaign

withdrawing their support, saying that they hadn't understood what was involved when they signed up.

The petition is waiting for the California Public Employment Relations Board to verify that all the signatures are valid. If the UAW has signatures from more than half of the eligible members, the union will automatically be adopted. If 30–50% of postdocs have signed, a system-wide election will be held. If less than 30% have signed, no action will be taken.

UAW representative Maureen

Boyd says the union had a "strong majority" when it filed the petition. At present, it isn't clear whether postdocs can withdraw their support.

"A lot of people didn't feel they got all the information they needed," says Gregory Potter, chair of the Postdoctoral Scholars Association of the University of California, San Francisco. He says that the UAW resisted invitations to public discussions on the topic.

A decision on the union is likely to be reached in the next few weeks. ■ Heidi Ledford

AIDS vaccine research becomes 'big science'

With no vaccine to show for more than 20 years of work, the HIV-vaccine community is being forced to radically change the way it works. Funding organizations are insisting on a 'big science' approach involving huge data-sharing collaborations. But AIDS researchers are divided over whether such a strategy will really speed progress towards a vaccine.

When the Bill & Melinda Gates Foundation announced a US\$300-million call for proposals in HIV vaccine research last year, it marked a key change for researchers used to chasing investigator-driven grants. Individual groups need not apply, said the foundation. Instead, it demanded that research groups organize themselves into bigger consortia, then jointly submit large, multimillion-dollar proposals with clear milestones and goals. The consortia had to agree to share data with each other, and to ensure that any patents won would allow vaccines to be made available to developing countries at low cost.

Researchers duly did as they were asked, and last month the Gates Foundation announced the winners of 16 grants, totalling \$287 million over five years. They will establish an international network of 11 vaccine-discovery consortia, plus 5 support and data-analysis facilities — bringing together 165 investigators from 19 countries in a single effort (see 'Some of the winners...').

The move comes a year after the US National Institutes of Health put some \$315 million over 7 years into a similar group of consortia to create the Center for HIV-AIDS Vaccine Immunology (CHAVI).

The idea of large consortia working on AIDS vaccine research was put forward in 2003 by a group of research funders. It was outlined earlier this year by the Global HIV Vaccine Enterprise, a virtual consortium aimed at accelerating HIV vaccine development that



J. DELAY/AP

A clutch of consortia will receive US\$287 million in an attempt to alleviate the plight of AIDS sufferers.

was endorsed by the G8 summit in July.

Instead of relying on researchers to form collaborations spontaneously, those with the money are now forcing it to happen, says Michel Kazatchkine, France's ambassador for AIDS and co-chair of the Enterprise coordination committee. Kazatchkine explains that instead of gathering and publishing their own data, researchers will be obliged to work with large, shared data sets, use standardized assays and methods, and farm out routine analyses to a series of centralized laboratories.

Much of this isn't controversial. Most AIDS scientists agree that standardized methods are

sorely needed, and there is broad support for the \$100 million that the Gates Foundation is providing for five central facilities to help structure and coordinate the field. At present, groups use different cohorts, assays and surrogate markers. This makes it all but impossible to compare results to find the most promising vaccine candidates for clinical trials. Individual researchers have little incentive to improve matters, as developing standards involves mundane work that does not yield prestigious publications.

More contentious are the 11 research grants to establish vaccine-discovery consortia — these are likely to be the talk of the corridors at

Some of the winners...

HIV poses a formidable challenge for vaccine developers: it mutates rapidly, attacks immune cells that might destroy it, and has an arsenal of tactics to avoid the body's defences, such as molecular shields that protect its surface from attack by antibodies.

AIDS researchers now believe that an effective vaccine will be two-pronged: as well as generating

antibodies, it will need to stimulate immune cells.

Of the 11 consortia selected, five will seek ways of generating effective antibodies when the immune system is exposed to HIV. Of these, a consortium led by Robin Weiss of University College London, who has won a \$25.3-million grant, will screen antibodies from humans and animals that seem naturally

active against HIV, and then work backwards to see which regions they target and design new vaccine candidates. And a \$19.4-million grant to a consortium led by Leo Stamatatos at Seattle Biomedical Research Institute will use computers to design molecules that might trigger antibodies against the virus.

The other six consortia will focus on stimulating a cellular response.

For example, Timothy Zamb of the International AIDS Vaccine Initiative has won \$23.7 million to try to modify other viruses to act as vectors for the vaccine. And a \$15.3-million project led by Giuseppe Pantaleo at the Centre Hospitalier Universitaire Vaudois in Lausanne, Switzerland, hopes to improve the ability of poxvirus, a known vaccine vector, to trigger a response to HIV. **D.B.**

A controversial choice?

The list of winners of the Gates Foundation grants has come under fire. Some researchers, including some winners, say in private that they are puzzled why certain projects were funded while others seen as more worthy were not.

This reflects the subjective nature of any review, other researchers say. But the secrecy of the process has fuelled speculation about bias. The review panel was anonymous, and researchers didn't receive the referee

reports typically given during grant applications.

"The names of the external reviewers were kept confidential to protect them from inappropriate pressures," says José Esparza, senior adviser on HIV vaccines at the Gates Foundation. He adds that scores and recommendations were not passed to applicants as it was a one-off competition with no possibility of resubmission, and that potential conflicts of interest were disclosed to panellists and foundation staff.

Each proposal was reviewed by 11 external experts, says Esparza. The foundation approved the ten highest-scoring projects, then selected four others from a combination of their scores and more subjective criteria, such as "maintaining a balance of projects; supporting novel ideas; and ensuring they weren't duplicating projects already supported by other funders". Two final grants were awarded to central facilities that would support the other projects. **D.B.**

to tell us to collaborate," he says.

Many researchers are concerned about the level of funding going to a mission-oriented approach versus investigator-driven science. "It's a difficult balance to get in vaccine discovery — it's not development, but research," says Weiss. "I don't think an AIDS vaccine has been held up because we didn't know how to collaborate. The limiting factor is a scientific breakthrough, a bright idea and new thinking."

Others argue that large mission-oriented grants could hold back the science. "You are going to get bad decisions," predicts one AIDS researcher who requested anonymity. He argues that the approach hands scientific choices to bureaucrats, and that the concentration of so much money in so few hands is a recipe for turf wars and nepotism — a criticism vocally expressed by US researchers left out of the CHAVI funding framework. He also questions whether large collaborations — which make sense in astronomy, where teams often study the same object at different wavelengths, or in tackling the vast amounts of data coming from the Human Genome Project — are as applicable to the experiments of vaccine research.

The task of making big science work for HIV falls to Adel Mahmoud, who will retire next month as president of Merck's Vaccine Division to become chief executive of the Global HIV Vaccine Enterprise.

Kazatchkine for one is a convert, arguing that the "incredible egotism" of individuals and institutions often creates communities that lack the vision to organize themselves productively: "If Mahmoud succeeds, he will demonstrate a new way of doing science, where collaboration becomes as important as competition, giving research a new image."

Declan Butler

See also pages 602 and 617.



Not everyone agrees with Bill Gates' idea of a mission-oriented approach to vaccine discovery.

next week's XVI International Aids Conference in Toronto, Canada. There are concerns about both the selection process (see 'A controversial choice?') and the wisdom of investing so much money in a big-science approach.

"We are having to learn a new culture and language relating to these milestones and directed research," says Robin Weiss of University College London, UK, who has won a \$25.3-million grant. His consortium involves 11 labs in at least 6 countries. Getting them to work together effectively will be an organizational challenge, Weiss says, although he adds that they already "feel like a family". But Weiss is one of several AIDS researchers who argue that the community already cooperates well. "I was a little surprised at the implication that we needed the Gates Foundation

ON THE RECORD

"Americans aren't gullible enough to believe that they came from a fish."

Creationist John Morris on a new Museum of Creationism in Kentucky.

"Today, with all the pollution, you cannot get cleaner water than melted ice-cap water."

Salik Haard promotes his new beer, made with water from Greenland's shrinking glaciers.

NUMBER CRUNCH

Scientists this week announced the longest-ever electronic tracking of a migrating animal, a diminutive seabird called the sooty shearwater.

65,000 km is the distance travelled by a sooty shearwater in 200 days.

5,000 km is the distance travelled each year by some caribou, the land animals that migrate the farthest.

800 g is the average weight of an adult sooty shearwater.

189 kg is the average weight of a caribou.

SCORECARD



Chemical analysis
Isotope ratio mass spectroscopy showed that testosterone found in the urine of Floyd Landis, winner of the Tour de France, did not have the same carbon isotope ratio as other hormones in the sample, showing it had a different, external source.



Climate porn
A report by the Institute for Public Policy Research, a UK think-tank, highlights the "quasi-religious register of doom, death, judgement, heaven and hell" in alarmist climate reporting that can become "secretly thrilling". This is not, the institute argues, a helpful way of framing things.

Sources: Associated Press, BBC News, Proc. Natl Acad. Sci. USA



G. RESZETER/ARDEA.COM

Views collide over fate of accelerator

Its parts have been dismembered, its roof is leaking, and a wall is missing. Now activists and scientists are squabbling over whether to completely raze the Bevatron — one of the most important particle accelerators ever built.

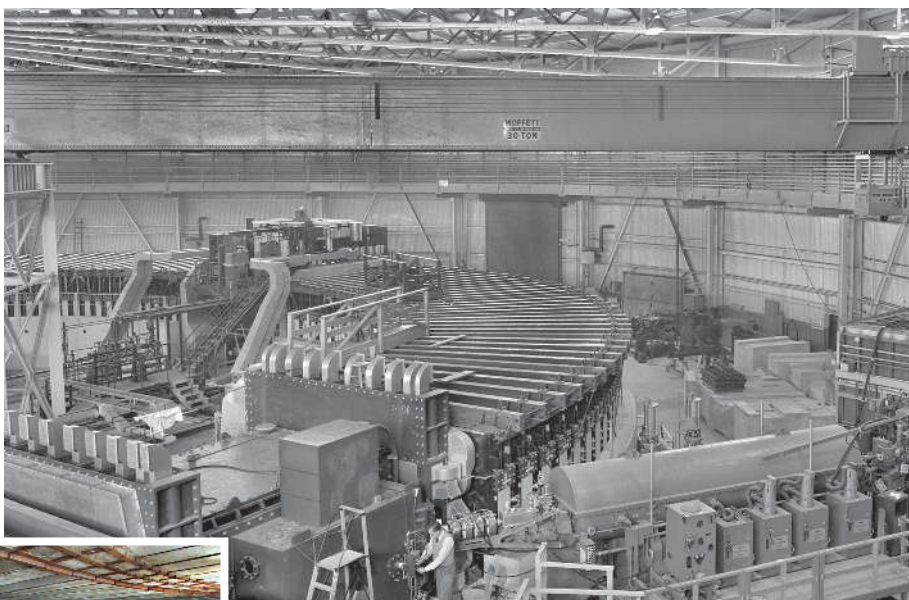
The remains of the Bevatron, which was decommissioned more than a decade ago, take up prime real estate on the Lawrence Berkeley National Laboratory's campus in Berkeley, California. Scientists at the lab want to tear it down to make way for fresh projects. But locals, many of whom oppose the demolition because of concerns about the possible release of contaminants, say they want to see it made into a museum.

On 3 August, the city council's Landmarks and Preservation Commission dealt a blow to those wanting landmark status for the accelerator by voting to recognize the Bevatron's legacy without protecting the building. Nevertheless, landmark advocates have vowed to continue fighting.

"It's truly a landmark, a very unique building," says Mark McDonald, who sits on the City of Berkeley's Peace and Justice Commission. "Somebody called it the world's largest yurt." The Bevatron, which from above looks not unlike a traditional Mongolian tent, began its life in 1954, smashing protons into a fixed target. A year after it opened for operation, physicists Emilio Segrè and Owen Chamberlain used it to discover the antiproton (see 'Bevatron's greatest hits'), for which they were awarded the Nobel prize in 1959. Over the next 40 years, other scientists used the Bevatron to study the violation of particle symmetry, as well as particle resonances, a phenomenon that eventually led to the quark model of matter.

But the machine had trouble keeping up in the fast-moving world of high-energy particle physics. Attempts to improve it were constrained by space restrictions — caused by the Berkeley Laboratory's placement on a steep hillside and close proximity to the city of Berkeley below. Over time it was gradually eclipsed by larger accelerators at other labs. In its last decades of operation, it was used to conduct biological research, including assessing the radiation risks associated with space flight. It was eventually shut down in 1993 due to budget shortfalls.

Since then, it has been taking up precious space on the Berkeley Lab's relatively small campus, according to Benedict Feinberg, a physicist there who served as the Bevatron's director for 1989–93. The laboratory recently got the \$90 million needed to remove the



LAWRENCE BERKELEY NAT'L LAB



Divided: physicists hope to reclaim the space but local groups want landmark status for the Bevatron.

machine, and is currently pursuing its demolition. "The value of the Bevatron is for the science that was done there," says Feinberg. "The building is nothing really special."

Feinberg declined to comment on the latest interest in preserving the building, but some lab officials quietly believe that it's the walls, not the accelerator, that are of real concern to local groups. The building was constructed using transite, a composite of asbestos and cement, which releases asbestos fibres when it is broken

up. Community members have expressed fears that razing the Bevatron would involve moving large amounts of loose asbestos through the city of Berkeley. Environmentalists also fear that lead and other contaminants from the building site could escape into the water table.

Pamela Sihvola, who co-chairs the Committee to Minimize Toxic Waste — a local watchdog group that monitors the laboratory — is among those pushing to preserve the Bevatron. "The EPA says that the best way to deal with lead and asbestos is to manage it in place," she says. "All we want the laboratory to do is consider a re-use alternative," she says. "It could be a living science museum and education centre." Chamberlain supported the idea, and even wrote a letter advocating the conversion of the Bevatron before his death in February, she notes.

McDonald adds that the centre would immortalize the Berkeley Lab's most famous work. "We haven't had a lot of Nobel prizes up there lately," he says. "There's nothing wrong with paying a tribute."

But Feinberg disagrees. With limited space, he argues, the Bevatron site is more valuable to lab researchers than the ageing machine it contains. "A fitting memorial," he says, "would be to re-use that site to do cutting-edge science."

The lab is now preparing the environmental-impact statement needed to proceed with demolition. Sihvola is gearing up for a fresh round of debate with the full city council: "Our attempts to preserve the building will continue," she vows.

Geoff Brumfiel

Bevatron's greatest hits

Most of the discoveries made in the Bevatron occurred early on in its 40-year history.

1955 Owen Chamberlain and Emilio Segrè use the Bevatron to find the antiproton, for which they win the Nobel prize in 1959.

1956 Bruce Cork's team discovers the antineutron using the accelerator.

1956 Tsung-Dao Lee and Chen Ning Yang suggest using the Bevatron to test 'parity violation', the idea that some laws of nature are not symmetric.

1960 Bogdan Maglich reports a unique particle called $Y^*(1385)$, which points to the existence of quarks.

School board regains evolutionist majority

Supporters of the teaching of evolution in US schools regained leverage in a key battleground state last week when Kansas voters kicked out two conservative members of the state board of education.

The primary election on 1 August means that the new board, once installed in January 2007, will have six members in favour of teaching evolution and four against. The board is expected to overturn teaching standards passed last November that labelled evolution “controversial”; these standards were backed by proponents of intelligent design — the notion that life was shaped by an intelligent designer.

The election is the latest swing in a seven-year tug-of-war. In 1999, the Kansas school board voted to essentially remove evolution from the curriculum, only to overturn that decision the next year when voters installed three moderates on the board. Conservatives regained the majority in 2004.

Supporters of evolution are claiming victory, even if it may only be a temporary one. “The intelligent-design movement is finding out that they can’t legislate science,” says Jack Krebs, president of Kansas Citizens for Science.

More drilling could benefit Louisiana’s wetlands

Oil, gas and healthy wetlands — maybe Louisiana can have it all.

On 1 August, the US Senate voted to open 8.3 million acres to oil and gas drilling — on top of the 40 million already open — on the Gulf of Mexico’s outer continental shelf. Half of the extra revenue would be shared among Alabama, Louisiana, Mississippi and Texas. A similar bill from the House of Representatives, passed in June, would open even more areas. The two bills must be reconciled if drilling is to begin.

Louisiana’s governor Kathleen Blanco is encouraging her state’s residents to vote for a measure that would allocate all the extra



Restoring wetlands could reduce the effect of hurricanes on cities such as New Orleans.

Museum goes ahead with online database

Archaeologists and anthropologists now have online access to a treasure trove of nearly 300,000 artefacts from Africa, Asia, the Americas and Oceania.

The new Quai Branly Museum in Paris, which opened in June, brings together collections of cultural items, textiles and art from France’s former colonies. The museum’s online database includes details of artefacts, such as the African mask pictured here.

Photographs, historical details and associated research for each artefact are available at the museum’s website, www.quaibranly.fr.



P. GRIES/B. DESCOINGS/MUSÉE DU QUAI BRANLY

revenue to coastal restoration and protection against hurricanes. Scientists agree that wetlands reduce the effect of hurricanes on land, and the plan has the support of such diverse groups as the American Petroleum Institute, the hunting organization Ducks Unlimited, and the conservation group Environmental Defense.

Japan lays out plans for a base on the Moon

Japan has announced a timetable for its plans to travel to the Moon. At a symposium on lunar exploration in Tokyo last week, an official at the Japan Aerospace Exploration Agency (JAXA) announced a deadline of 2020 for sending astronauts to the Moon, and 2030 for constructing a base there.

But the latest plans could cost ¥3 trillion (US\$26 billion), and sceptics are already speaking up. “To go the Moon would cost so much that I don’t think anyone could afford to, except possibly America,” says Bruno Gardini of the European Space Agency in Noordwijk, the Netherlands.

A more modest proposal comes from South Africa, where the cabinet in July approved plans to bring the country’s space activities under the umbrella of a new space agency. Funding has yet to be agreed, but government officials say they hope the agency will be established within a year.

Senators set sights on stem-cell legislation

Congressional backers of human embryonic stem-cell research have a target date for reintroducing legislation to expand support for such research. In January 2007, a bipartisan group of senators plans to bring up the bill as one of its first orders of business with the newly elected

Congress, according to Senator Tom Harkin (Democrat, Iowa).

President George W. Bush vetoed the bill last month (see *Nature* 442, 335; 2006), but a two-thirds majority in both houses of Congress could override his decision. “The president’s veto is not the last word,” said Harkin, who pushed for the legislation. He claims that embryonic stem-cell research will be a key issue in November’s elections.

Also last week, the liberal think-tank Center for American Progress released a report arguing that state support for embryonic stem-cell research is insufficient, and that federal funding is needed.

Politicians think locally and act globally on climate

The United States is willing to tackle climate change, at least at a local level. Two high-profile partnerships formed last week aim to tackle greenhouse gases regionally — but with international help.

Former president Bill Clinton has joined forces with mayors from 22 cities on six continents, including London, Los Angeles and Mexico City. The Clinton Climate Initiative will encourage its partners to share resources to reduce the amount of greenhouse gases in the atmosphere.

And British prime minister Tony Blair last week signed up with California governor Arnold Schwarzenegger, whose state is a leader in innovative technologies to cut pollution. The deal is only a statement of intent, but it could bring California into the active UK carbon-trading market.

Correction

The News Feature “A little goes a long way” (*Nature* 442, 351–352; 2006) incorrectly suggested that Steven Ley carries out chemical reactions at -195°C , the temperature of liquid nitrogen. Experiments are in fact performed at -78°C , the temperature of solid carbon dioxide in acetone.

BUSINESS

Silent running: the race to the clinic

A technique for turning genes off has sparked a flurry of biotech investment. **Erika Check** investigates.

Silence is golden — at least, that's what two US biotechnology companies are hoping. They aim to reap the rewards of bringing a method for silencing genes into the clinic, where they claim it could lead to treatments for everything from viral diseases to cancer.

The technique, known as RNA interference or RNAi, could offer a safe and effective way of turning off a gene. And, although it is early days, many see it as a likely moneyspinner for the biotechnology industry.

"It is a while away from being an impactful sector," says Sapna Srivastava, an analyst at Morgan Stanley in New York. "But it has the potential to be really big."

Two well-funded companies are leading the way for now: Alnylam Pharmaceuticals of Cambridge, Massachusetts, and Sirna Therapeutics of Boulder, Colorado. These two firms are jousting over intellectual-property rights, and each has been forging corporate alliances in a bid to capitalize on RNAi's potential.

Researchers are watching keenly as this power struggle unfolds. Some are anxious to see whether commercial pressure and the lust for profit will stifle the spread of knowledge in this burgeoning field, or lead to failures that could derail it — as has happened to an extent in gene therapy.

RNAi is a natural process, thought to combat viruses or gene disruption in some organisms. It was first discovered in plants, but was soon identified in organisms such as nematode worms and fruitflies. In 2000, a team of four scientists including Nobel laureate Phillip Sharp of the Massachusetts Institute of Technology in Cambridge, revealed the molecular mechanism behind RNAi in fruitflies¹. A year later, one of the co-authors, Thomas Tuschl, who was then at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany, came up with a technique for using RNAi in mammals². As a result, the four scientists founded Alnylam in 2002.

The company's access to Tuschl's key patents on using RNAi, coupled with the founding team's impressive scientific credentials, gave it a huge jumpstart. Since starting up, Alnylam has raised \$50 million in venture capital and \$100 million from public stock offerings. Its stock has risen from \$6 a share to nearly \$13 a share, and it has raked in



Howard Robin, chief executive of Sirna Therapeutics, celebrates his company's success on the Nasdaq.

some \$100 million through deals with Merck, Medtronic and Novartis.

The Novartis deal, signed last September, was seen in the industry as a watershed. Worth up to \$700 million if Alnylam successfully develops therapeutic leads against diseases such as pandemic flu, the deal was by far the largest in the field at that time. "This is one of the most significant discovery alliances in biotech history," says John Maraganore, Alnylam's chief executive.

The rise of Alnylam's rival was decidedly less meteoric. Originally called Ribozyme Pharmaceuticals, Sirna was founded in 1993 by Nobel laureate Thomas Cech, then a biochemist at the University of Colorado at Boulder and now head of the Howard Hughes Medical Institute in Chevy Chase, Maryland.

The company focused on developing RNA enzymes, called ribozymes, for use as human therapies. It went public in 1996 but almost got kicked off the Nasdaq in 2002 when its stock price slumped after disappointing trials of ribozymes to treat cancer and hepatitis C.

A last-ditch survival bid saw the company's scientists turn to RNAi. The firm soon raised \$48 million in venture capital, changed its name to Sirna and retooled itself as an RNAi outfit. Its market capitalization has since soared from

\$4 million to more than \$375 million. Echoing Alnylam's success, Sirna this May announced a deal with GlaxoSmithKline that could yield up to \$700 million if it develops successful therapies for four respiratory diseases.

The involvement of major drug firms in RNAi attests to the optimism that surrounds the technology. "Pharmaceutical companies see it as their third class of drug targets, after small molecules and proteins," says Srivastava.

Despite its promise, RNAi therapy still faces one significant hurdle: drug delivery. The technique works by introducing a short section of RNA into the body that can then interfere with — and switch off — the gene in question. But RNA is rapidly chewed up by the immune system, so it is hard to make sure it reaches its genetic targets inside cells. Both Sirna and Alnylam are racing to address this issue and to crack the conundrum of 'systemic delivery' — in which the drug will find its way to the specific cells in question.

Sirna performs extensive chemical modifications to its drug molecules and has devised a delivery system that uses lipid nanoparticles. In effect, these are protective fatty casings that help the RNA to cross cellular membranes. The company plans to test its system in clinical trials as early as this autumn.

Alnylam, meanwhile, has also tested a variety of delivery packages and this March announced promising results for one of them. Packaging the

"Success will rest heavily on which method for inducing gene silencing works best — and which firm ends up holding the rights."

RNA inside fatty globules, Alnylam researchers successfully injected the drug into monkeys and silenced the expression of a gene that makes cholesterol. The resulting paper³ is the first to show that systemic RNAi might work in primates.

The drug-delivery problem has meant that RNAi's initial forays into the clinic have been aimed at diseases affecting the eye. This is because the RNA molecules can be delivered to the eye without coming into contact with the immune cells that would destroy them.

The first treatment to enter clinical trials was not from either Alnylam or Sirna, but from Acuity Pharmaceuticals of Philadelphia, Pennsylvania. Its treatment for macular degeneration, a common cause of adult blindness, began safety trials in October 2004. It has since completed phase II efficacy trials, and claims that the drug is effective — although it has yet to release the full data. Sirna is not far behind with its treatment for macular degeneration, which will enter phase II trials by the end of this year.

Alnylam, on the other hand, abandoned its therapy for the condition, correctly predicting that Genentech of South San Francisco, California, would quickly secure approval for its antibody treatment for the disease (see *Nature* 422, 123; 2006). Alnylam's most advanced programme uses direct delivery of RNA to the lungs by inhalation. It has completed two phase I studies for treatment of the respiratory syncytial virus, and plans to initiate another lung study later this year aimed at pandemic flu.

But Alnylam and Sirna are almost as heavily focused on claiming intellectual property as they are on drug discovery. Their respective prospects will rest heavily on which method for inducing gene silencing works best — and which firm ends up holding the rights to it. Alnylam says its exclusive licence to one of Tuschl's patents gives it the edge over Sirna. But Sirna claims its recent innovations bypass Tuschl's patents completely.

So far, investors are more impressed with Alnylam's position, says Srivastava. "We spoke with Novartis, and they said they signed their partnership with Alnylam mostly because of intellectual property," she says.

With their broad interests in treatments for different diseases, both Sirna and Alnylam dream of following in the footsteps of California's biotech powerhouses Genentech, and Amgen of Thousand Oaks.

"We think we have a good shot at becoming a major, biotechnology-based pharmaceutical company, given this technology," says Howard Robin, Sirna's chief executive. ■

1. Zamore, P. D., Tuschl, T., Sharp, P. A. & Bartel, D. P. *Cell* 101, 25–33 (2000).

2. Elbashir, S. M. et al. *Nature* 411, 494–498 (2001).

3. Zimmermann, T. S. et al. *Nature* 441, 111–114 (2006).

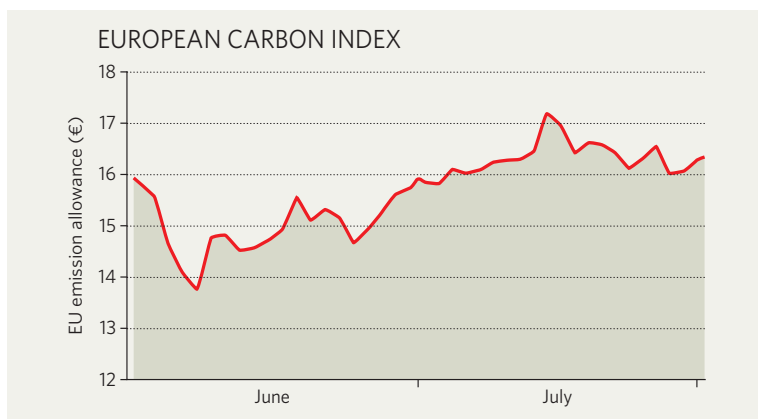
IN BRIEF

TIME FOR PLAN B The US Food and Drug Administration (FDA) says it will consider permitting pharmacies to sell the morning-after contraceptive, Plan B. Critics have attacked the regulator for failing to allow such sales on non-scientific grounds, three years after its advisory panels said that they should go ahead. The day before Andrew von Eschenbach, who has been nominated by President Bush to serve as the FDA's commissioner, faced a Senate confirmation hearing, the agency announced that it wants to meet the Plan B's manufacturer, Barr Pharmaceuticals, to discuss terms for approval. Von Eschenbach's failure to approve the treatment while he was acting commissioner was a focal point of the hearing — but some senators said afterwards they'd still block his confirmation.

CHIPS ON THE TABLE The semiconductor industry is sitting on a US\$2-billion surplus of microchips, mainly because of overproduction at Intel, the largest chip producer, says California-based market-research firm iSuppli. The number of chips worldwide almost doubled between the first and second quarters of 2006, the firm reports. But it says that because the glut of chips is largely down to miscalculations at just one manufacturer, it is unlikely to cause the industry any major problems.

MERCK COURT WIN A court in California has rejected a claim by a 71-year-old man that his heart attack was caused by taking Vioxx, the painkiller withdrawn from sale by Merck in September 2004. This latest ruling means that the drug firm has now won five Vioxx cases out of eight, leading some analysts to conclude that its strategy of fighting the cases one-by-one is bearing fruit. Merck says there are 14,200 cases pending in the United States over alleged side effects of the drug (see *Nature* 440, 277; 2006).

MARKET WATCH



The European market for emissions trading has recovered its balance after turbulence earlier in the spring, when uncertainty about the market's future saw allowances lose half of their value.

Options to emit a tonne of carbon dioxide are traded on five European exchanges, including the European Energy Exchange in Leipzig, Germany, charted here. The price rose from a low of about €10 (US\$13) per tonne in early May to a high of €18 on 18 July, before settling in a band of between €15 and €17. Although heatwaves throughout Europe have driven the demand for fossil energy, emission allowances are significantly cheaper than during the cooler summer of 2005, when the price peaked at almost €30.

This year's lower price is no surprise to market analysts. "Everybody knows by now that governments have over-

allocated emission rights to industry," says Stefan Kleeberg of Climate Change Consulting in Frankfurt, Germany. Increased sales bring down the price every time it approaches the €17 margin, which speculative traders believe is about the maximum value the options can currently achieve, he says.

The volume of transactions is down by half compared with its peaks last year, but analysts say that it is likely to rise again after the summer break.

Carbon dioxide emissions are expected to rise too, says Guy Turner, director of New Carbon Finance in London, as high oil and gas prices are forcing an increased use of coal in electricity generation. But analysts doubt whether prices on the emission allowances will go much higher, as most industries own plenty of surplus options. ■

Quirin Schiermeier

STAYING THE COURSE

Some feared that widespread use of AIDS treatments in Africa would encourage drug resistance, with globally disastrous consequences. But there's no crisis yet, reports **Erika Check**.

The argument was simple, even brutal: there was no point giving anti-AIDS drugs to patients in Africa, because they wouldn't be able to stick with their treatment plans. Worse, the resulting rise of resistant strains of HIV could jeopardize lives elsewhere.

This counsel of despair was first expressed five years ago, after some 17 million people in sub-Saharan Africa had already died of AIDS. It emerged while researchers and politicians debated whether to roll out antiretroviral medications to the millions who needed them in developing countries. Aid agencies and governments found the argument unconvincing, even distasteful, and pressed on. As a result, the epidemic has now reached an important turning point. For the first time, more people are taking anti-AIDS medications in poor countries than in rich ones. They number more than a million — three times the number at the end of 2001 (ref. 1). So, is a crisis looming?

Not so as you'd notice. Despite some important exceptions, there is strong evidence that patients in the poorest countries are just as conscientious about taking their medications as those in the West, and widespread resistance has not yet emerged. But scientists warn that the next step — expanding drug programmes to reach a greater proportion of those who need treatment — will be much more challenging than anything faced so far.

Getting patients to comply with their drug regimens has always been a major issue in the treatment of AIDS. The medications work by blocking HIV during various parts of its life cycle in human cells. But the virus has a notoriously error-prone process of replicating itself, so it can easily mutate to become resistant to treatment. Taking combinations of medicines suppresses the virus and cuts the chances of new, treatment-resistant mutations. But if a patient doesn't take all of his or her drugs on a tight schedule, resistant strains can emerge, and be passed on to other people.

This is already a serious and growing problem in the developed world. A 2005 UK study showed, for example, that, depending on the drug, up to 20% of new patients were infected with strains that were already resistant². Problems with adherence were a factor, but such strains mostly arose in patients taking drugs that were less effective than the combination therapies that later became available.

Back in 2001, Africa was seen as a worst-case scenario when it came to adherence to combination therapies. Even if drugs could be made affordable, poor nations wouldn't necessarily be able to coordinate effective mechanisms to



S. MAINA/AFP/GETTY

These antiretroviral drugs prolong the lives of people with AIDS. But not all can gain access to them.

deliver them. Most patients didn't have access to regular medical services. And it was unclear whether they would trust Western medicine enough to stick with a long-term course of medicine with toxic side effects.

Some had more controversial concerns. There were warnings that the very existence of drug programmes in Africa could breed complacency and encourage risky behaviour such as unprotected sex, thus increasing the spread of any resistant strains that arose (see 'Do drugs lead to risky behaviour?'). And most famously, Andrew Natsios, then the director of the US Agency for International Development, argued against giving antiretroviral drugs to patients in Africa, because these patients "don't know what Western time is" and so would be inca-

pable of staying on their treatment plans. Ask patients to stick to a complex regime, Natsios argued, and they "do not know what you are talking about".

Such criticisms might sound prejudiced, if not downright racist, but those advocating them pointed out that if drug programmes in Africa failed, the whole world could come to regret it. "If treatments are not adhered to consistently and correctly, there could be disastrous consequences both for individuals on antiretroviral therapy, and for the HIV epidemic as a whole," wrote social scientists Danielle Popp and Jeffrey Fisher of the University of Connecticut, Storrs, in a 2002 article³. "Developing countries could become a veritable 'petri dish' for new treatment-resistant strains," they warned. Such

strains could spread around the globe, rendering antiretroviral drugs useless elsewhere.

Researchers involved in rolling out treatments at the time say they were acutely aware of the dangers. "We were all going into this with our eyes open to resistance issues," says Marc Wainberg, who will co-chair the XVI International AIDS Conference, to be held in Toronto over 13–18 August. Some suggested modelling HIV programmes on principles similar to the directly observed therapy (DOTS) strategy for tuberculosis, in which patients take their medication under the eye of an observer⁴. But this isn't so practical for HIV, because patients have to take drugs every day for the rest of their lives. So researchers who set up some of the earliest sites for antiretroviral treatment in sub-Saharan Africa designed other measures to help patients stay straight.

For example, in 2002, the Botswanan government became the first in Africa to promise AIDS treatment to everyone who needed it. The programme built on a pilot project supported by the government and international funders, including drug companies, universities and the Bill & Melinda Gates Foundation. Patients were asked to nominate a family member or friend as an 'adherence assistant', to help them monitor side effects and stick to their treatment courses.

It seemed to work. Scientists found that nearly 80% of the first 153 patients treated on the programme knocked the virus down to undetectable levels after 48 weeks on drugs — a good sign that they had stuck to their medication⁵.

Richard Marlink, director of care and treatment at the Elizabeth Glaser Pediatric AIDS Foundation in Santa Monica, California, and one of the scientists behind the project, sees such results as a stinging rebuttal of those who doubted Africans' ability to stick to complex schedules. "All the indications from studies we've done or anecdotally show that you can obtain levels of adherence you'd attain anywhere else in the world," he says. "And there

are some indications that they are better."

Other studies support the idea that patients in even the most challenging environments can manage AIDS treatment regimens. In South Africa, the non-profit group Médecins Sans Frontières has been working since 1999 with local government to deliver HIV drugs to the residents of the Khayelitsha township. In 2004, doctors with the programme reported that of the first 287 patients treated, 70% had undetectable levels of virus after two years of treatment⁶. The results echoed earlier findings from a widely publicized study performed in Cape Town, which found that 66% of patients still on medication after 48 weeks had undetectable virus loads⁷. Other studies have found good, although not exceptional, results in places such as Senegal and Uganda^{8,9}.

One forthcoming analysis even suggests that, so far, African patients have a better combined record than those in rich nations. Edward Mills and an international group of investigators compared the results of studies on populations in Africa and North America. They were surprised to find that whereas the studies on North Americans reported that 55% of the

populations covered stuck to their treatment regimens, 77% of the populations in African studies did so¹⁰. "The common expectation was that poverty was a risk factor for incomplete adherence," says David Bangsberg, director of the Epidemiology and Prevention Interventions Center at San Francisco General Hospital and senior author on the study. "That's wrong, in retrospect."

Bangsberg says his team's analysis suggests that other factors may be more important than poverty — such as lack of social support, addiction to drugs or alcohol, bad relationships with the healthcare system and homelessness, which in Western settings have all been shown to make people more likely to stop taking their drugs. "It isn't poverty that's a risk factor for adherence," says Bangsberg. "It's some of the things that go with poverty in North America

"Poverty was expected to be a risk factor. That was wrong."
— David Bangsberg



Tactics: promoting AIDS awareness is as much a part of the battle as providing the right drugs.

and western Europe that aren't necessarily part of poverty in southern and western Africa."

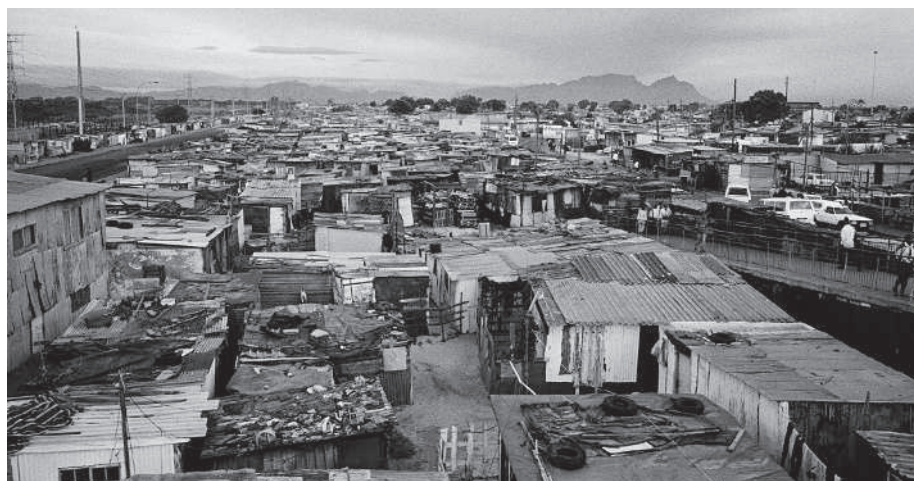
Others aren't sure things are quite so rosy. Last year, Christopher Gill and his colleagues at the Boston University School of Public Health, Massachusetts, released a study in which they surveyed conference abstracts for unpublished reports on adherence in Africa. One abstract reported that only 48% of patients in a Côte d'Ivoire study stuck to their treatments, and although another found 68% adherence in a population in Cameroon, that rate declined over time¹¹.

Bad bias

Overall, the authors found that adherence rates were much more variable than those published in the literature, which has led them to ask whether bad news just isn't getting reported. "This would seem to be a classic example of publication bias," says Gill. "People are blowing their horns when things are great, and keeping quiet when things aren't."

Bangsberg, however, argues that the low adherence rates in these unpublished studies often weren't the fault of the patients. He contends that adherence rates in one Senegalese study were poor because patients had to buy their own drugs. So if they ran out of money, they couldn't stick to the regimen. And he says that in the Cameroonian study, the drug supply was interrupted by factors beyond the patients' control¹². To Bangsberg, these results don't show that patients can't stick to their drugs, but demonstrate the pressing need to strengthen drug-distribution networks.

Indeed, scientists say that fixing such structural problems will be key to keeping adherence rates high, as drug programmes now



Model town: residents of Khayelitsha, in South Africa, are successfully keeping to their drug regimes.



move beyond pilot projects into the wider population.

Despite progress so far, the majority of patients in sub-Saharan Africa who need antiretroviral drugs still aren't getting them. Botswana is perhaps furthest ahead, but according to conservative estimates, only about half the patients who need treatment there are on it. The rest of the region lags far behind; from South Africa, where about one-third of patients who need drugs are on them, to Zimbabwe, where just 5–10% of those who need treatment can get it. "Sadly, with all the effort and money being put into treatment access, there is still a holocaust going on," says David Katzenstein, drug-resistance specialist at Stanford Univer-

sity, California. He estimates that, globally, 5–10% of the people who will die from AIDS during the next year are actually being treated.

Expanding programmes to reach more people looks likely to be much tougher than anything faced so far. Patients in clinical trials receive many incentives to stay on their treatments, and are carefully followed by specialist teams. But as treatment programmes make the transition from pilot sites to general public-health programmes, not everyone will get such close attention. And the patients who started treatment in the first roll-out programmes may start experiencing problems as the long-term toxicity of some of the drugs kick in. "It's still early days in the roll-out," says Dan Kuritzkes, director of AIDS research at the Brigham and Women's Hospital in Boston. "I'm sure we'll enter a phase two, which is going to be more challenging than the early phase."

A grand plan

Because of this, scientists say it's important to make plans now to monitor the emergence of large-scale drug resistance. In rich nations, doctors are recommended to genotype individual patients' viruses before beginning therapy, to check whether they are resistant to any particular drugs. But such tests cost hundreds of dollars, and it's not feasible to deploy them for every patient starting therapy in sub-Saharan Africa. So a World Health Organization effort called the Global HIV Drug Resistance Surveillance Network has set recommendations for measures countries should take as they scale up treatment — including tracking an easily accessible group of newly infected patients, such as young women pregnant for the first time, and a cohort of patients who stop responding to drugs.

If patients do fail first-line therapy, there's little

that can be done, because second-line treatments are still too expensive to be widely available in Africa. So as well as trying to increase access to those second-line drugs, the main concern is keeping everyone adherent. But support costs money, and poor countries will have to make trade-offs, Gill warns. "Do you want to have a good programme for fewer people, or a mediocre programme for a lot of people?" he asks. "Someone has to make that decision."

To stand any chance of success there must be full commitment to providing a robust drug supply and to supporting patients, say scientists, and that means moving on from the idea that Africans aren't capable of taking their drugs properly. "The hard work ahead of us is to go beyond pilot projects, to taking the health of Africa seriously," says Marlink. "It's time to put these concerns about adherence to bed." ■

Erika Check covers the biomedical sciences for Nature.

1. UNAIDS/WHO AIDS Epidemic Update: Special Report on HIV Prevention (UNAIDS, Geneva, 2005).
2. UK Group on Transmitted HIV Drug Resistance *Br. Med. J.* **331**, 1368–1371 (2005).
3. Popp, D. & Fisher, J. D. *AIDS* **16**, 676–678 (2002).
4. Harries, A. D. et al. *Lancet* **358**, 410–414 (2001).
5. Wester, C. W. et al. *J. Acq. Immun. Def. Syn.* **40**, 336–343 (2005).
6. Coetzee, D. et al. *AIDS* **18**, 887–895 (2004).
7. Orrell, C., Bangsberg, D. R., Badri, M. & Wood, R. *AIDS* **17**, 1369–1375 (2003).
8. Laurent, C. *AIDS* **16**, 1363–1370 (2002).
9. Weidle, P. J. et al. *Lancet* **360**, 34–40 (2002).
10. Mills, E. et al. *J. Am. Med. Assoc.* **296**, 679–690 (2006).
11. Gill, C. J., Hamer, D. H., Simon, J. L., Thea, D. M. & Sabin, L. L. *AIDS* **19**, 1243–1249 (2005).
12. Bangsberg, D. R., Ware, N. & Simoni, J. M. *AIDS* **20**, 140–141 (2006).
13. Morin, S. F. et al. *AIDS Behav.* **7**, 353–362 (2003).
14. Jha, P. et al. Commission on Macroeconomics and Health Working Paper WGS:2 (2001).
15. Moatti, J.-P. et al. *AIDS* **17** (Suppl. 3), S69–S77 (2003).
16. Bateganya, M. et al. *AIDS Patient Care STDs* **19**, 760–768 (2005).
17. Bunnell, R. et al. *AIDS* **20**, 85–92 (2006).

Do drugs lead to risky behaviour?

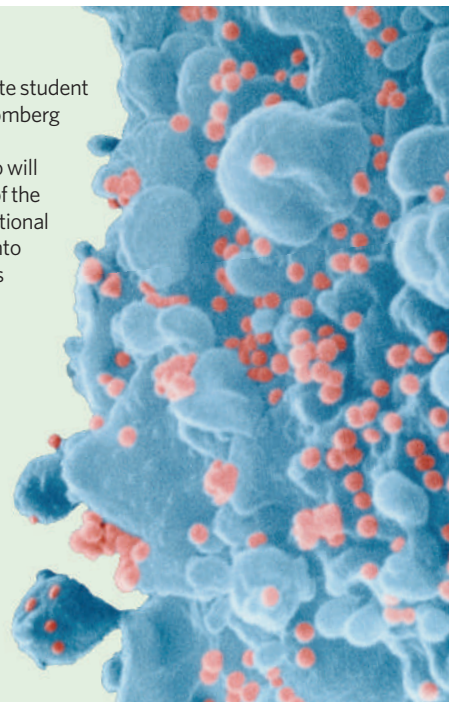
Thanks to antiretroviral therapy, AIDS is no longer an immediate death sentence in rich nations. That's heartening progress, but there is a downside. Social scientists believe it has also contributed to the phenomenon of 'disinhibition' — increasing rates of high-risk behaviour, for example the rise in unprotected sex that has been observed in some communities of gay men¹³.

Before drug therapy was rolled out in Africa, some were concerned that it might cause an increase in risky behaviour there, too. For example, Frank Plummer, director-general of the Centre for Infectious Disease Prevention and Control in Canada, cited evidence that condom use among prostitutes in Nairobi went down when quack cures for AIDS were reported¹⁴.

But so far, the evidence says that access to drugs isn't increasing risky sex — and it may actually help people to be safer. A 2003 study of 711 people with HIV (pictured) in Côte d'Ivoire found that patients not treated with antiretrovirals were more likely to report risky sexual behaviours than those who received treatment¹⁵. In another study, Ugandans who were treated with antiretroviral drugs were more likely to tell their partners about their HIV-positive status, use condoms with those partners, and seek treatment for other sexually transmitted infections¹⁶. And when antiretrovirals were provided along with counselling, in a separate study, the rate of risky behaviour in the group dropped by 70% (ref. 17).

"These studies all show the same trend, and it's encouraging," says

Caitlin Kennedy, a graduate student at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, who will present a meta-analysis of the studies at the XVI International AIDS Conference in Toronto next week. But the studies done so far aren't high rigorous, and the treatment programmes are still in progress. Researchers say it is too early to draw any firm conclusions. They point out that disinhibition is unlikely to occur in Africa unless treatment becomes so widely available that most HIV-positive people no longer die from AIDS — a point that is sadly far from being reached





PUTTING THE CARBON BACK

One way to keep carbon dioxide out of the atmosphere is to put it back in the ground. In the first of two News Features on carbon sequestration, **Quirin Schiermeier** asks when the world's coal-fired power plants will start storing away their carbon. In the second, **Emma Marris** joins the enthusiasts who think that enriching Earth's soils with charcoal can help avert global warming, reduce the need for fertilizers, and greatly increase the size of turnips.

The hundred billion tonne challenge

Ketzin, a dozy village of 4,000 people west of Berlin, hardly looks like a vision of the future. Nestled in the Havel countryside — an idyllic mix of rivers and forests — it has a small tourist industry and, as is typical for such parts of eastern Germany, a not-so-small unemployment problem. “There’s no news at the moment”, says the community’s website.

But there could be news soon. In 2004, Ketzin was chosen as the site of mainland Europe’s first large-scale carbon storage demonstration project. By the end of the year, drilling will start at a former gas storage facility on the edge of town. In the next two years, some 60,000 tonnes of carbon dioxide (roughly the annual carbon dioxide output of 40,000 cars) will be injected into an aquifer of salty water 700 metres beneath the surface.

The Ketzin project will test the ‘storage’ part of carbon capture and storage (CCS), a strategy designed to allow energy to be generated from fossil fuels without the carbon dioxide produced in the process ending up in the atmosphere.

Little more than a fringe idea five years ago,

CCS was singled out at the 2005 G8 summit as a technology that could make a difference to climate change. Experts see it as a central part of any strategy for maintaining the generation of energy at today’s levels: as Vassilios Kougionas, a European Commission officer in charge of clean-coal initiatives and international energy relations, puts it, “Without CCS there is no point in continuing with fossil fuels.”

And yet, for all this enthusiasm, there is a distinct lack of urgency in government approaches. The countries most interested in CCS have, at best, preliminary plans for it; most haven’t even got that far. Meanwhile, the number of power stations whose carbon dioxide is neither captured nor stored is rising inexorably, as is the atmospheric level of the gas.

To some observers, this represents a failure not of science or technology, but of will. “It does require quite substantial research and technology efforts to make CCS better and more efficient — but we could still start today if desired,” says Hans Ziock, a physicist at Los Alamos National Laboratory in New Mexico, who has worked on carbon dioxide capture

technologies for more than a decade. David Hawkins, director of the National Resources Defense Council’s Climate Center in Washington DC, goes further: “Global efforts are completely out of scale with what is needed.”

Act now or pay later

The International Energy Agency (IEA) predicts that by 2030, global energy demand will grow by 1.7% a year. Although contributions from nuclear and renewable sources may increase, the IEA predicts that 85% of the rise in demand will be met by greater use of fossil fuels. This means the overall capacity of coal-fired power plants will have to double in that time, from 1,100 gigawatts to 2,200 gigawatts¹. Taking into account the existing plants that will be shut down, the world is looking at 1,400 gigawatts’ worth of new coal plants (see graphic on page 623). This doubling of capacity is the greatest expansion of power generation in the planet’s history.

Since 1750, humanity’s burning of coal has released about 150 billion tonnes (gigatonnes) of carbon into the atmosphere. During their lifetimes, the new generation of plants will

release 140 gigatonnes. Meanwhile, climate scientists are arguing that carbon dioxide levels should not be allowed to get much higher than 550 parts per million (p.p.m.); the current level is 380 p.p.m., which compares with 280 p.p.m. in the eighteenth century. Some argue that the ceiling needs to be a lot lower.

Robert Socolow is a physicist and co-principal investigator of the Carbon Mitigation Initiative at Princeton University in New Jersey. He calculates that if carbon dioxide levels are to be kept in the desired range, then humanity needs to avoid about a third of the emissions expected in the next 50 years. That means finding a way to not release 175 gigatonnes of carbon. "If we get going now," he says, "the job will be less than half as difficult. If we don't it means we're running a very costly strategy."

The good news is that, in principle, such vast amounts of gas could indeed be tackled by CCS. Deep aquifers in the world's sedimentary basins have a total storage capacity estimated at between 1,000 and 10,000 gigatonnes². Pumping carbon dioxide into them is a straightforward matter. Oil companies already pump carbon dioxide into petroleum reservoirs on a routine basis as a way of flushing out the hydrocarbons. And experience with oil shows that such reservoirs can keep their contents stored away for geological lengths of time, points out Günter Borm, director of geo-engineering at Germany's National Research Centre for Geosciences in Potsdam and coordinator for the European Union-funded Ketzin project.

Not all reservoirs are as well adapted to CCS as a 100-million-year-old oilfield might be. Some may leak, and in some there might be a risk of sudden, catastrophic releases of gas. In low-lying land, such leakages could suffocate people because carbon dioxide is heavier than air, so will fill up valleys and basins. Under the sea, gas-filled reservoirs could potentially trigger landslides and thus tsunamis. But provided storage sites are chosen carefully, designed for safe operation, and properly monitored, the risks are manageable, says Lynn Orr, director of the Global Climate and Energy Project at Stanford University in California.

At Ketzin, scientists will keep track of any undesired chemical interactions between carbon dioxide and minerals, which could in principle dissolve the 'cap-rock' that seals a storage site, or contaminate drinkable ground water. Scientists monitoring a smaller storage project, the Frio Brine Pilot Experiment in Texas, recently reported that the injection of carbon dioxide had made the brine 1,500 metres down substantially more acid³; such acidic brine could potentially eat through the surrounding rock and escape into higher aquifers.

But Hawkins counsels against too much pilot-project research of this type. He thinks that each reservoir will have unique geologi-

cal characteristics that will need to be assessed *in situ*. He also points out that even if some reservoirs leak, it still makes more sense to use them, and thus spread out emissions over time rather than do nothing.

Everyone agrees that large-scale carbon dioxide storage would be a gargantuan technical feat. Locking away 250 million tonnes of carbon per year — equivalent to 4% of annual global emissions — would require an injection of 25 million to 35 million barrels per day, depending on compression density. That's equivalent to about a third of the flow of oil currently coming from reservoirs.

According to Orr, if the infrastructure used to pump carbon dioxide into the ground was roughly the same size as the infrastructure currently deployed to bring oil to the surface, it could deal with about a seventh of the world's production of fossil-fuel-generated carbon. That is less than half the amount produced at power stations and large factories — the sources for which CCS is best suited.

Ground work

"It's a big enterprise," says Socolow. But pipeline building and well-drilling are mature and remarkably inexpensive technologies, and running costs would be extremely low. Experts calculate that setting up storage facilities, each capturing several million tonnes of carbon per year, for the carbon dioxide produced from hundreds or thousands of plants, might cost as much as \$80 billion. In a world set to invest \$16 trillion in energy by 2030, \$80 billion is not an unthinkable amount; the cost of deep disposal for Britain's nuclear waste has recently been estimated as £11.3 billion (\$21 billion). What's more, the figure might come down as the technology matures and economies of scale cut in.

As yet, industrial-scale CCS activities are limited to just three sites — in Norway, Canada and Algeria — and to megatonnes rather than gigatonnes of carbon. Since it began in 1996, the Norwegian project has pumped around 10 million tonnes of carbon dioxide 1,000 metres beneath the North Sea bed into the Utsira Sand formation; the carbon dioxide is a contaminant in natural gas from the Sleipner West field. In the Canadian and Algerian projects, the gas being stored away is also being used to enhance the productivity of oilfields, which covers some of the costs. All these sites have been continuously monitored for possible leakages, but none seems to have occurred.

In total, 11 or so full-scale CCS projects are planned, or have been proposed, in the United States, Canada, Britain, Germany, Norway and Australia. But given present trends, experts think it unlikely that CCS will be employed at any substantial level before 2030, by which time the 1,400-gigawatts' worth of power stations will already have been built. For CCS to

"If we get going now, the job will be less than half as difficult." — Robert Socolow





Flow reversal: Statoil has pumped around 10 million tonnes of carbon dioxide, separated from the Sleipner West natural gas field, beneath the North Sea bed.

make a difference, the mechanisms it requires must be available to a substantial fraction of those plants after they have been built.

Conventional coal-burning technology, which will probably be used to produce 1,200 of the 1,400 gigawatts, produces flue gases that are about 14% carbon dioxide. If the carbon dioxide is to be captured and stored it must first be 'scrubbed' from the gas stream, typically by running the gases through an amine solution. This takes up the carbon dioxide and then, when heated, releases it in pure form. The problem with using this technique is that the equipment needed not only costs money — it also takes up a lot of space. Fitting such equipment into a plant not designed for it is expensive.

The main alternative to the conventional plant is the integrated gasification combined cycle (IGCC). The capital costs for IGCC plants are 20% greater than for conventional plants. But they have environmental benefits — among which is being cheaper to kit out for CCS.

In IGCC plants, the fuel — coal, fuel oil or biomass — is introduced into a hot gasifier along with oxygen and steam. This produces a fuel gas consisting mainly of carbon monoxide and hydrogen. The carbon monoxide then goes through a second 'shift' reaction with steam, making carbon dioxide and more hydrogen. The carbon dioxide can be relatively easily separated at this point.

Four IGCC plants are currently in operation — two in the United States, one in the Netherlands and one in Spain. Although more expensive, and less profitable, than conventional plants, they have very low emissions of sulphur dioxide and other pollutants. Carbon dioxide capture from plants such as these would be significantly easier and cheaper than from conventional plants.

Making it pay

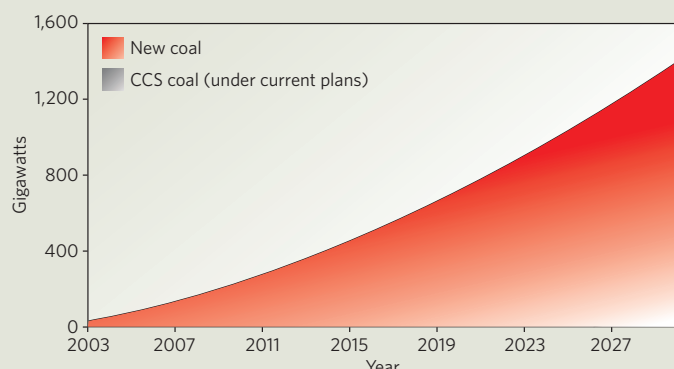
At the moment, the efficiency of IGCC plants is about 40%, which is roughly the same as that for good conventional plants. CCS would drop the efficiency of both sorts of plant to about 30%. But IGCC plants use two thermodynamic cycles; the hydrogen from the gasifier and the shift reaction drives a gas turbine while the heat from that turbine and the gasifier drives a separate steam turbine. Having more than one cycle means that, in principle, it should be possible to push the overall efficiency much higher, to the point where, even when paying the energy penalty associated with CCS, the plants would still be competitive. "First comes efficiency, then CCS," says Jacek Podkanski, a senior energy technology specialist with the IEA.

The US 'FutureGen' initiative, a \$1-billion public-private partnership to design, build, and operate a coal-fuelled zero-emissions power plant by 2015, aims to prove the technical and economic feasibility of a commercial-scale IGCC plant fitted with CCS. In Europe,

STATOIL

SOURCE: IEA, D. HAWKINS

PREDICTED RISE IN CAPACITY OF COAL-FIRED PLANTS



the German energy company RWE Power has recently announced that it will invest €1 billion (US\$1.3 billion) in the construction of a 450-megawatt IGCC plant in Germany. The plant, fully equipped for CCS, could become operational in 2014. Other companies, such as Swedish Vattenfall, are also experimenting with ways to get more efficiency out of traditional plants, and with 'oxyfuel' plants in which the coal is burned in pure oxygen, producing a flue gas that is richer in carbon dioxide and so better suited to CCS.

"We're by no means proceeding as though there were all the time in the world, but new technologies need time," says Johannes Ewers, head of new power-plant technologies at RWE. "We do know that if we want to burn coal we have to reduce carbon dioxide."

In theory, the EU's emissions trading system should be providing incentives for such developments by putting a price on emissions. Such incentives have been shown to work: the sequestration of carbon dioxide from the Sleipner platform in Norway is the oil and gas company Statoil's response to the country's \$50

costs less than \$20 on the European exchange, and analysts doubt that the price is likely to increase much in the future. "The price of emitting carbon dioxide must rise, otherwise it just won't work," says Borm.

"CCS is perfectly doable at modest costs," says Jon Gibbins, a CCS expert at Imperial College London and project manager with the UK Carbon Capture and Storage Consortium. "But politically we're unfortunately in a very chaotic situation where it's hard to look forward for more than one or two years. If you want to tackle climate change you should do it aggressively or not bother at all."

The road ahead

For Gibbins, the best way forward is to ensure that every new plant is 'capture-ready' — that is, designed in such a way that CCS mechanisms can be relatively easily incorporated at a later date. The UK government is considering requiring that all new coal-fired plants

per tonne of carbon dioxide tax on emission-intensive industries. The cost of capturing, transporting and storing a tonne of carbon dioxide is estimated by the Intergovernmental Panel on Climate Change to be between \$20 and \$70, which is why the Statoil sequestering project makes sense. But emitting a tonne of carbon dioxide currently

are capture-ready. The idea is also part of an EU–Chinese 'memorandum of understanding' on near-zero-emissions power generation technology, and was highlighted in the 2005 G8 action plan.

But some fear that capture-readiness will just provide a cheap way of doing a small amount to cut emissions. "The term 'capture-ready' is pretty meaningless, because its definition includes subsequent installation of unidentified equipment," says Hawkins.

To him, it would make more sense for governments to set up well-defined performance standards for power-generating facilities, such as a maximum carbon dioxide output allowed per unit of electricity. Such standards could become gradually more strenuous as better technologies become available. Another possibility, Socolow points out, would be simply to subsidize the technology. He thinks 2–3 cents per kilowatt-hour would make CCS a profitable route for a new coal-fired plant using today's technologies.

That is only a little more than the 1.8-cent subsidy the United States currently gives producers of wind-generated electricity.

No matter which technological options and economic or regulatory incentives are used, CCS will cost society money. "Electricity will clearly become more expensive," says Ewers. How expensive depends on technology choices

and geographical circumstances, but rough estimates suggest that the average production costs of electricity will double, from 4 to 7 cents per kilowatt-hour. This might be bad news for coal producers, as, all other things being equal, it would make the fuel they sell much less desirable, tilting the market towards natural gas. But it would not necessarily be a terrible burden for consumers, says Socolow. As the costs of distribution and transmission are hardly affected, he calculates that the retail cost of electricity would increase by just 20%.

These costs, says Gibbins, are trifles compared to the long-term benefits of allowing companies to go on generating power with proven technology but without adding to the greenhouse effect. He also points to a peculiar long-term advantage. If fossil fuels are left unburned, they could be used in the future. If burned with CCS, they are gone for good, with the environment unchanged. "It doesn't make a lot of difference in which form carbon stays in the ground," he says. "But with carbon dioxide rather than desirable fossil fuels in the ground you have a very different path ahead."

Quirin Schiermeier is *Nature's* German correspondent.

1. International Energy Agency *World Energy Outlook 2004* www.iea.org/Textbase/publications/free_new_Desc.asp?PUBS_ID=1266
2. IPCC Special Report: *Carbon Dioxide Capture and Storage* http://arch.rivm.nl/env/int/ipcc/pages_media/SRCCS-final/ccsspm.pdf (2005).
3. Kharaka, Y. K. et al. *Geology* **34**, 577–580 (2006).

See Editorial, page 601

M. DUFFY/ALAMY



Mechanisms to store carbon will draw on well-established technologies already used by the oil industry.

J. LEHMANN



Drop of the black stuff: *terra preta* contrasts strongly with normal soil in colour (left) and produces much more vigorous crops (below).

Black is the new green

In 1879, the explorer Herbert Smith regaled the readers of *Scribner's Monthly* with tales of the Amazon, covering everything from the tastiness of tapirs to the extraordinary fecundity of the sugar plantations. "The cane-field itself," he wrote of one rum-making operation, "is a splendid sight; the stalks ten feet high in many places, and as big as one's wrist." The secret, he went on, was "the rich *terra preta*, 'black land', the best on the Amazons. It is a fine, dark loam, a foot, and often two feet thick."

Last month, the heirs to Smith's enthusiasm met in a hotel room in Philadelphia, Pennsylvania, during the World Congress of Soil Science. Their agenda was to take *terra preta* from the annals of history and the backwaters

of the Amazon into the twenty-first century world of carbon sequestration and biofuels.

They want to follow what the green revolution did for the developing world's plants with a black revolution for the world's soils. They are aware that this is a tough sell, not least because hardly anyone outside the room has heard of their product. But that does not dissuade them: more than one eye in the room had a distinctly evangelical gleam.

The soil scientists, archaeologists, geographers, agronomists, and anthropologists who study *terra preta* now agree that the Amazon's dark earths, *terra preta do índio*, were made by the river basin's original human residents, who were much more numerous than formerly supposed. The darkest patches correspond to the

middens of settlements and are cluttered with crescents of broken pottery. The larger patches were once agricultural areas that the farmers enriched with charred trash of all sorts. Some soils are thought to be 7,000 years old. Compared with the surrounding soil, *terra preta* can contain three times as much phosphorus and nitrogen. And as its colour indicates, it contains far more carbon. In samples taken in Brazil by William Woods, an expert in abandoned settlements at the University of Kansas in Lawrence, the *terra preta* was up to 9% carbon, compared with 0.5% for plain soil from places nearby¹.

From Smith's time onwards, the sparse scholarly discussion of *terra preta* was focused mainly on the question of whether 'savages' could have been so clever as to enhance their land's fertility. But Woods' comprehensive bibliography on the subject now doubles in size every decade. About 40% of the papers it contains were published in the past six years.

Loam ranger

The main stimulus for this interest was the work of Wim Sombroek, who died in 2003 and is still mourned in the field. Sombroek was born in the Netherlands in 1934 and lived through the Dutch famine of 1944 — the *Hongerwinter*. His family kept body and soul together with the help of a small plot of land made rich and dark by generations of laborious fertilization. Sombroek's father improved the land in part by strewing it with the ash and cinders from their home. When, in the 50s, Sombroek came across *terra preta* in the Amazon, it reminded him of that life-giving 'plaggen' soil, and he more or less fell in love. His 1966 book *Amazon Soils* began the scientific study of *terra preta*.

Since then trial after trial with crop after crop has shown how remarkably fertile the *terra preta*



EPRIDA/UGA BIO-CONVERSION CENTER

is. Bruno Glaser, of the University of Bayreuth, Germany, a sometime collaborator of Sombroek's, estimates that productivity of crops in *terra preta* is twice that of crops grown in nearby soils². But it is easier to measure the effect than explain it through detailed analysis.

Everyone agrees that the explanation lies in large part with the char (or biochar) that gives the soil its darkness. This char is made when organic matter smoulders in an oxygen-poor environment, rather than burns. The particles of char produced this way are somehow able to gather up nutrients and water that might otherwise be washed down below the reach of roots. They become homes for populations of microorganisms that turn the soil into that spongy, fragrant, dark material that gardeners everywhere love to plunge their hands into. The char is not the only good stuff in *terra preta* — additions such as excrement and bone probably play a role too — but it is the most important factor.

Leaving aside the subtleties of how char particles improve fertility, the sheer amount of carbon they can stash away is phenomenal. In 1992, Sombroek published his first work on the potential of *terra preta* as a tool for carbon sequestration³. According to Glaser's research, a hectare of metre-deep *terra preta* can contain 250 tonnes of carbon, as opposed to 100 tonnes in unimproved soils from similar parent material. The extra carbon is not just in the char — it's also in the organic carbon and enhanced bacterial biomass that the char sustains.

Ground control

That difference of 150 tonnes is greater than the amount of carbon in a hectare's worth of plants. That means turning unimproved soil into *terra preta* can store away more carbon than growing a tropical forest from scratch on the same piece of land, before you even start to make use of its enhanced fertility. Johannes Lehmann of Cornell University in Ithaca, New

York, has studied with Glaser and worked with Sombroek. He estimates that by the end of this century *terra preta* schemes, in combination with biofuel programmes, could store up to 9.5 billion tonnes of carbon a year — more than is emitted by all today's fossil-fuel use⁴.

Mud pack

The year before he died, Sombroek helped to round up like-minded colleagues into the Terra Preta Nova group, which looks at the usefulness of using char in large-scale farming and as a carbon sink. The group was well represented at the Philadelphia meeting, although Glaser was not there. Their aim is to move beyond the small projects in which many of them are involved and find ways of integrating char into agribusiness. After all, wherever there is biomass that farmers want to get rid of and that no one can eat, char is a possibility. That means there are a lot of possibilities.

One problem is that there is a new competitor for farm waste. Plant are largely made up of cellulose, indigestible material in cell walls. Recent technological advances make it likely that quite a lot of that cellulose might be turned into biofuel. At the moment, ethanol is made from corn in the United States and from sugar in Brazil; if it were made directly from cellulose, producers could work with a wider range of cheaper biomass. Given the choice of turning waste material into fuel or into charcoal, farmers might be expected to go for fuel, especially if that is the way that policy-makers are pushing them: US President George W. Bush promised \$150 million for work on cellulosic ethanol in his 2006 state of the union speech.

But Lehmann and his colleagues don't see biofuel as an alternative to char — they see the two developing hand in hand. Take the work of Danny Day, the founder of Eprida. This "for-profit social-purpose enterprise" in Athens, Georgia, builds contraptions that farmers can use to turn farm waste into biofuel while making char. Farm waste (or a crop designed for biofuel use) is smouldered — pyrolysed, in the jargon — and this process gives off volatile organic molecules, which can be used as a basis for biodiesel or turned into hydrogen with the help of steam. After the pyrolysis, half of the starting material will be used up and half will be char. That can then be put back on the fields, where it will sequester carbon and help grow the next crop.

Negative thinking

The remarkable thing about this process is that, even after the fuel has been burned, more carbon dioxide is removed from the atmosphere than is put back. Traditional biofuels claim to be 'carbon neutral', because the carbon dioxide assimilated by the growing biomass makes up for the carbon dioxide given off by the burning of the fuel. But as Lehmann points out, systems such as Day's go one step further: "They are the only way to make a fuel that is actually carbon negative".



C. STEINER

Slow burn: the idea of using charcoal to sequester carbon may take a while to catch on.

Day's pilot plant processes 10 to 25 kg of Georgia peanut hulls and pine pellets every hour. From 100 kg of biomass, the group gets 46 kg of carbon — half as char — and around 5 kg of hydrogen, enough to go 500 kilometres in a hydrogen-fuel-cell car (not that there are many around yet). Originally, Day was mostly interested in making biofuel; the char was just something he threw out, or used to make carbon filters. Then he discovered that his employees were reaping the culinary benefits of the enormous turnips that had sprung up on the piles of char lying around at the plant. Combining this char with ammonium bicarbonate, made using steam-recovered hydrogen, creates a soil additive that is now one of his process's selling points; the ammonium bicarbonate is a nitrogen-based fertilizer.

"We don't maximize for hydrogen; we don't maximize for biodiesel; we don't maximize for char," says Day. "By being a little bit inefficient with each, we approximate nature and get a completely efficient cycle." Robert Brown, an engineer at Iowa State University in Ames, has a \$1.8-million grant from the United States Department of Agriculture (USDA) to fine-tune similar technology, although being in Iowa, he uses corn stalks not peanut hulls. "We are trying an integrated approach: we are trying to evaluate the agronomic value, the sequestration value, the economic value, the engineering," he says.

Brown thinks a 250-hectare farm on a char-and-ammonium-nitrate system can sequester 1,900 tonnes of carbon a year. A crude calcu-





lation on that basis suggests the US corn crop could sequester 250 million tonnes of carbon a year. At the moment, no one knows how long this could go on; no one has yet found a ceiling for char addition.

Stephen Joseph of Biomass Energy Services and Technology in New South Wales has built a number of char-producing machines in Australia that work at fairly large scales (the models have grown from an original 'Piglet' through a larger 'Daisy' to a positively bullish 'El Toro'). Joseph looks for companies with a waste problem such as a paper mill with spare scraps or a dairy with old bedding and manure, and then integrates char production into the business so that the heat produced in pyrolysis is used where the firm needs it.

So far, Joseph's company is being brought in to solve waste-management problems, but he hopes the value of the char will become a selling point in itself. For that to happen, however, he needs some help. His machines can be tuned to make char of various sorts: different sized particles with different sized pores and different amounts of other elements. Which is the best? It's a question he asks in Philadelphia, and one of the things that Brown's research in Iowa aims to find out.

The right protocol

Such technical unknowns are not the only obstacles on the road to a black revolution. One problem is that the purported benefits of char do not slot easily into the framework of the Kyoto Protocol, an international agreement to reduce carbon emissions. Lehmann hopes to see the process get going under the aegis of the protocol's Clean Development Mechanism, in which rich countries sponsor green projects in poor countries and get credit for the reduced emissions. To this end, he is amassing evidence that modern char techniques can actually keep the carbon involved locked up for centuries. His Cornell colleague John Gaunt is working on ways to present the technique as the sort of 'change in practice' that could count as a tradeable carbon-emission reduction of the sort allowed under Kyoto.

Then there are your risk-averse farmers. They haven't heard of char. And they aren't going to buy it — let alone buy a strangely named machine to make it — unless they know it will make them money. It is no good pitching it to them with a mouthful of scientific caveats about not knowing the right kind of char for each type of soil or exactly how it works. You have to be able to sell specific benefits and real attractions. "A lot of farmers are environmentalists," says John Kimble, a USDA man who has just retired from the National Soil Survey Center in Lincoln, Nebraska. "But they look at the bottom line, as we all do."

After the afternoon coffee break in Philadelphia, Kimble takes the podium and the



Father of the field: Wim Sombroek championed the study of the Amazon's dark soils.

wind out of everyone's sails. He is sympathetic to the terra pretans goals — indeed he was a good friend of Sombroek's — but that doesn't stop him asking hard questions. "Can you do this in a no-till way?" is one tricky query. Kimble and many others have been pushing no-till farming, which basically means doing without ploughs, as a partial solution to erosion, pesticide run-off and fuel costs. The idea is that the less you mess with the soil, the less its components separate and wander away. But biochar is light and fine, like the black grit left in a barbecue. If you don't physically insert it into the soil, it might just blow away.

Everyone listens politely. But while watching their responses, it was hard not to worry that the same enthusiasm that has brought them together might also trap them in a cul de sac. They obviously respect economics and pragmatic requirements. But these are not people principally moved by practical politics or bottom lines; they are people moved by ideals. They start from the basis that the answer lies in the soil, more or less whatever the question is, and can't quite understand why this isn't self-evident to everyone else. Faced, for example,

with the suggestion that all corn matter be turned into ethanol, they tend simply to say "Well it could be — but we hope, of course, it will go into the soil." They know they ought to be marketing *terra preta* as a resource, or a policy instrument; but they can't stop

seeing it as a wonder.

Policy is not always, or even often, dictated by pure rationality. Perhaps *terra preta's* compelling history and rich, earthy smell will go to the heads of that diffuse band of policy-makers who hand out the cash. The enthusiasts need to be more down to earth; but the policy people might benefit from getting their hands dirty. ■

Emma Marris is a Washington correspondent for *Nature*.

"This is the only way to make a fuel that is actually carbon negative."
— Johannes Lehmann

1. Woods, W. I. & McCann, J. M. in *Yearbook Conf. Latin Am. Geogr.* Vol. 25 (ed. Caviedes, C.) 7–14. (Univ. Texas, Austin, 1999).
2. Glaser, B. *Phil. Trans. R. Soc.* (in the press).
3. Sombroek, W. G. *Interciencia* **17**, 269–272 (1992).
4. Lehmann, J., Gaunt, J. & Rondon, M. *Mitigation Adapt. Strateg. Global Change* **11**, 403–427 doi:10.1007/s11027-005-9006-5 (2006).

Scientists are well placed to speak up for biodiversity

SIR — As reported in your News story “Should conservation biologists push policies?” (*Nature* 442, 13; 2006), academics are reluctant to demand the implementation of schemes to halt and reverse the biodiversity crisis. Conservation scientists seem to fear that taking a position would mean appearing biased, and that this would be better left to others.

Using science to conserve biodiversity does not require a biased opinion: it simply requires belief in the results. There is ample evidence showing that ecosystem, species and genetic diversity are critical for mankind's well-being, and that biodiversity is declining.

Conservation scientists should follow the path of Richard Doll, who revealed that smoking causes lung cancer (R. Doll and A. B. Hill. *Br. Med. J.* 2, 739–748; 1950) and then became active in advocating changes in public-health policies — without undermining his academic credibility. If Doll, and later his colleague Richard Peto, had adopted a ‘back-seat’ attitude and not lobbied against the tobacco industry, many more lives would be claimed by lung cancer today than is the case.

Guillaume Chapron

Carnivoreconservation.org,
address supplied

Authors were clear about hockey-stick uncertainties

SIR — Your News story “Academy affirms hockey-stick graph” (*Nature* 441, 1032; 2006) states that the US National Academy of Sciences (NAS) panel “concluded that systematic uncertainties in climate records from before 1600 were not communicated as clearly as they could have been”. This conclusion is not stated in the NAS report itself, but formed part of the remarks made by Gerald North, the NAS committee chair, at the press conference announcing the report.

The name of our paper is “Northern Hemisphere temperatures during the past millennium: inferences, uncertainties, and limitations” (*Geophys. Res. Lett.* 26, 759–762; 1999). In the abstract, we state: “We focus not just on the reconstructions, but on the uncertainties therein, and important caveats” and note that “expanded uncertainties prevent decisive conclusions for the period prior to AD 1400”. We conclude by stating: “more widespread high-resolution data are needed before more confident conclusions can be reached.” It is hard to imagine how much more explicit we could have been about the uncertainties in the reconstruction; indeed, that was the point of the article!

The subsequent confusion about uncertainties was the result of poor communication by others, who used our temperature reconstruction without the reservations that we had stated clearly.
Raymond S. Bradley*, **Malcolm K. Hughes†**,
Michael E. Mann‡

*Department of Geosciences,
University of Massachusetts, Amherst,
Massachusetts 01003, USA

†Laboratory of Tree-Ring Research,
University of Arizona, Tucson,
Arizona 85721, USA

‡Department of Meteorology,
Pennsylvania State University,
University Park, Pennsylvania 16802, USA

Systems biology could help us harness useful microbes

SIR — I would like to add my strong support to your Editorial “Infection biology” (*Nature* 441, 255–256; 2006) praising the innovative approaches to infection biology described in the US National Academies report *Treating Infectious Diseases in a Microbial World*. We need to move on from viewing infection as the rather exotic province of a handful of specialized ‘virulence genes’ carried by a small group of pathogenic organisms, to an appreciation of our true complicity in the microbial world.

The initial encounter of a pathogen with the host immune system involves a bewilderingly complex interchange of signals between the two organisms that sets in train a multiplicity of respective adaptive responses. In addition to immune cells and the pathogen itself, our personal complement of 1.5 kg of intestinal microflora is likely to be a major arbiter in this encounter. Commensal microbes are important in maintaining the networks regulating the innate immune response: they are certainly important in distributing genes mediating resistance to antibiotics, and they may also regulate the way in which we metabolize treatment drugs.

Systems biology provides a framework in which to address this complexity. Moving towards the rational design of novel antimicrobials and novel vaccines — particularly novel antimicrobials and vaccines that can be applied in synergistic combinations — involves digestion of huge amounts of information generated by increasingly high-throughput approaches to microbiology, immunology, genetics and population biology. To make sense of this requires a combination of traditional biology skills with the expertise of mathematicians and computer scientists.

In addition to improving our weapons in the ‘war’ against selected pathogens, a deeper understanding of the ways in which we interact with microbes is crucial in

combating a wide range of diseases and disorders. The population of microbes that has a vested interest in a healthy human population is very much larger than that which seeks our destruction. Systems biology may help us turn this fact to our advantage.

Douglas Young

Centre for Integrative Systems Biology,
Imperial College London, Flowers Building,
Armstrong Road, London SW7 2AZ, UK
www.imperial.ac.uk/cisbic

Sticking points in the push for change

SIR — Your Editorial “Reaching a tipping point” (*Nature* 441, 785; 2006) discusses in the context of climate change the concept of a ‘tipping point’, from Malcolm Gladwell’s book of the same name: that extraordinary change is possible if only a few key pre-conditions come together appropriately.

Gladwell’s proposition embeds an older, specious belief: that directed change of the social or political world is easy, perhaps even easier than changing the physical world. But compare, for example, the building of the first atomic bomb with the elimination of racism or poverty.

When social change does occur, it is seldom as dramatic as the kinds of change described by tipping points. There is considerable continuity in social systems because of the resilience of institutions, and because, as articulated by Anthony Giddens in *Modernity and Self-Identity: Self and Society in the Late Modern Age* (Stanford Univ. Press, Stanford, 1991), of the basic human need for ‘ontological security’ — a firm belief in the continuity of social and material environments, and in the endurance of social patterns and individual self-identity.

Even if there were widespread agreement on the need for a specific social change, via a tipping point, the procedures for accomplishing this are entirely unclear. Gladwell writes about the considerable parallels between the idea of tipping points and the diffusion of innovations, but neglects the fact that more than 90% of innovations fail.

As quoted in your Editorial, Gladwell’s argument runs: “Look at the world around you. It may seem like an immovable, implacable place. It is not. With the slightest push — in just the right place — it can be tipped.”

The conclusion is valid, as far as it goes, but is highly misleading because of its rarity. The usual reality is a generally immovable world where a lot of pushing has little effect.

Eugene A. Rosa

Thomas S. Foley Institute for Public Policy and Public Service, Washington State University,
Pullman, Washington 99164-4020, USA

COMMENTARY

Fair payment or undue inducement?

Women who donate their eggs for stem-cell research should be compensated in the same way as other healthy research volunteers, argues Insoo Hyun.

Women who undergo hormonal induction to generate oocytes for embryonic stem-cell research can receive financial compensation for their direct expenses only, according to ethical guidelines issued by the US National Academy of Sciences, the California Institute for Regenerative Medicine and the state of Massachusetts. This narrow payment restriction has emerged as a conventional position among US policy-makers, but I will argue that it is ethically unwarranted and unfair. Women who donate oocytes for stem-cell research should also be compensated for their time, effort and inconvenience.

Despite the existing guidelines, there is still a chance to expand the scope of permissible payments to oocyte providers. Last month, the International Society for Stem Cell Research (ISSCR), for which I am chair of the Subcommittee on Human Materials Procurement, released draft international guidelines that do not expressly limit remuneration to women's direct expenses. It is seeking comments from ISSCR members and the international community on its website (www.isscr.org/scientists/guidelines.cfm) before finalizing its recommendations. Consequently, I hope that the issue of oocyte provider compensation will remain open for deliberation in research locales where the compensation level has not yet been set.

Why should women be compensated for bearing the burdens of hormonal induction for stem-cell research? Oocyte providers fall under the category of healthy research volunteers and should be treated as such. It is common practice for healthy volunteers to be compensated for the inconvenience of participating in biomedical research^{1,2}. The US National Institutes of Health (NIH) regularly compensates healthy volunteers for their participation in protocols conducted through its Clinical Center in Bethesda, Maryland³. The NIH maintains that it is ethically acceptable to compensate such volunteers for their troubles, especially if the protocol offers them little or no direct benefits⁴. And the Institutional Review Board Guidebook of the US Department of Health and Human Services (DHHS) states that compensation is permissible as long as local ethical review bodies ensure that volunteers are fairly recruited,



Retrieving eggs from volunteers involves minor surgery, which can be painful and carries health risks.

adequately informed and paid according to reasonable community standards for research participation².

These common practices belie the common assumption that a ban on oocyte provider compensation is the correct ethical default position. Oocyte procurement is demanding of women's time and energy, especially when done in an ethically rigorous and reasonably safe manner. Women may devote as many as 40 cumulative hours to provide their oocytes for research (C. Racowsky, personal communication). This includes pre-enrollment interviews for comprehensive informed consent, research protocol education, consultations with physicians, nurses and social workers, and medical and laboratory testing.

Painful procedure

Once enrolled, women receive painful hormone injections daily for up to 3 weeks to produce controlled ovarian hyperstimulation. Oocytes are retrieved through a minor surgical procedure that involves painkillers, sedatives and, if necessary, general anaesthesia. Foreseeable risks, although rare, include bleeding, infection, cramping, mood swings, pregnancy and complications associated with general anaesthesia. Hormonal induction also

carries a 0.05–5% chance of severe ovarian hyperstimulation syndrome, a serious medical condition that, if left unattended, can cause kidney failure, infertility and, in extremely rare cases, death⁵. There may also be long-term health risks (see special report on page 606).

The crux of the issue is this: if it is ethically and legally permissible for women to offer their oocytes for stem-cell research, and if it is acceptable to compensate healthy volunteers for their time, effort and inconvenience when undergoing comparable invasive procedures for research, such as donation of bone marrow or a bronchoscopy, then there is a strong, presumptive reason to compensate healthy women who provide oocytes for basic research. Those who want to limit remuneration to women's direct expenses must rebut this presumption. To succeed, they must do more than warn that compensating oocyte providers might introduce ethically undesirable consequences, for both sides of the debate readily acknowledge there are dangers. Rather, they must argue that there is no practical way to avoid these negative consequences for women.

What are the potential negative consequences of compensating oocyte providers? One worry is that the offer of compensation

C. PRIEST/SPL



Stem-cell banks rely on oocyte donations to further embryonic stem-cell research.

could undermine women's informed consent by creating an undue inducement to provide oocytes for research. Undue inducements involve the offer of a reward so irresistible that it blinds individuals to the disadvantages of participating in the proposed activity⁶.

The implication that compensation will impair women from adequately considering their risks is far from inevitable. The possibility of undue inducement is not a new problem in research ethics, and there are strategies that ethical review bodies can use to prevent it. For instance, the DHHS has recommended that local review bodies set compensation levels to reasonable and accepted standards for research participation within their communities, and that they monitor researchers' recruitment and informed-consent procedures².

Eggs for sale?

In addition to these institutional safeguards, the informed consent process itself could be used to turn the threat of undue inducement on its head. The disclosure of compensation presents an excellent teaching opportunity to underscore the demands and risks of oocyte procurement and to communicate the message that oocyte providers are essentially working to advance the goals of the principal investigator. Compensation by itself does not introduce the risk of undue inducement. Women's capacity to give voluntary and informed consent can be impaired even in the absence of compensation if consent procedures are not properly followed, as was the case with several women donors who contributed to Woo Suk Hwang's research in South Korea⁷. (I was involved in revising the informed-consent procedures for Hwang's research but sadly these were not followed⁸.)

Another worry is that compensation could have the unintended effect of enticing socio-economically disadvantaged women to volunteer as oocyte providers. It is unclear whether this concern is mainly about undue inducement, which we have just addressed, or about the exploitation of vulnerable populations. If the latter, then it is worth noting that, for

decades, ethical review bodies have been responsible for scrutinizing researchers' recruitment strategies to ensure that vulnerable populations are not unjustly enlisted. Oocyte procurement for stem-cell research should not be held to a lesser standard.

It is also empirically doubtful that the offer of compensation alone would greatly increase women's willingness to participate. A recent study of hypertension patients suggested that increased compensation levels do not significantly alter research volunteers' perception of risk, and that raising payment levels is less influential for poorer volunteers than for wealthier ones⁹. These preliminary data, coupled with the supervisory bodies' usual responsibilities, should help ease concerns about undue inducement and exploitation of the poor.

Nevertheless, some opponents contend that oocyte provider compensation is wrong because it introduces an element of commodification. Commodification occurs when market value is placed on something and it is treated as a product to be bought and sold. In this case, the commodification of human eggs is not the relevant worry, because compensation would be for women's time, effort and inconvenience, not the number or the quality of their oocytes. But such compensation can still be perceived as wrong because it puts a price on women's reproductive labour. Underlying this position is the assumption that women's reproductive labour is too sacred to be bought and sold.

This commodification argument is unconvincing for two reasons. First, the doctrine of informed consent requires that the decision to participate in research ultimately resides with the volunteer, who is free to decide for herself whether her reproductive labour is too sacred. Second, as with all research involving healthy volunteers, local ethical review bodies should set the appropriate level of compensation, not the free market, not oocyte providers and not stem-cell researchers. Although it is true that

women are often paid handsomely by US fertility clinics to provide oocytes for clinical purposes, the largely unregulated nature of this activity is a historical accident, and ethical review bodies need not follow suit. Making oocyte provider compensation subject to rigorous review by these bodies allows women's compensation levels to be regulated, capped and scrutinized to a degree unmatched by most fertility clinics — and far from the conditions found in a free market.

Others oppose compensation because of a belief that women should be motivated by pure altruism to provide their oocytes for research. Altruistic donation is widely considered ideal in clinical contexts, such as live organ donation, where the act carries clear therapeutic benefit. However, this paradigm is unsuitable for oocyte providers in stem-cell research because this scientific field is still in the early stages of basic research⁹. Because the burdens of oocyte procurement are so high and the immediate scientific benefits of using human oocytes in stem-cell research uncertain at best, it is inappropriate to ask women to undergo hormonal induction out of pure altruism.

Cultural norms around the world have demanded for centuries that women sacrifice their bodies to advance others' interests,

"It is inappropriate to ask women to undergo hormonal induction out of pure altruism."

usually with inadequate recognition of their endeavours. Stem-cell research should not emulate this historical trend. Compensation offers a reasonable way to acknowledge women's efforts by rightly embracing oocyte providers as healthy research volunteers. In locales where oocyte donation for stem-cell research is permitted and where healthy

research volunteers are usually paid, the question for policymakers is not whether to allow oocyte provider compensation. The question rather is how best to determine women's compensation levels and to monitor their recruitment and informed-consent processes. ■

Insoo Hyun is in the Department of Bioethics at Case Western Reserve University School of Medicine, Cleveland, Ohio, USA.

1. Food and Drug Administration IRB Information Sheets www.fda.gov/oc/ohrt/irbs/toc4.html#payment (2001).
2. US Department of Health and Human Services IRB Guidebook www.hhs.gov/ohrp/irb/irb_chapter3.htm (1993).
3. The Clinical Research Volunteer Program for Healthy Volunteers at the National Institutes of Health; www.cc.nih.gov/ccc/prc/healthyvols.html (21 June 2006).
4. NIH Remuneration of Research Subjects in the Intramural Research Program www.nihtraining.com/ohsr/site/info/sheet20.html (2000).
5. Delvigne, A. & Rozenberg, S. *Hum. Reprod. Update* **8**, 559–577 (2002).
6. Emanuel, E. J. L. *Med. Ethics* **32**, 100–105 (2004).
7. Steinbrook, R. N. *Engl. J. Med.* **354**, 324–326 (2006).
8. McGee, G. A. *J. Bioeth.* **6**, W33 (2006).
9. Halpern, S. D. *et al. Arch. Intern. Med.* **164**, 801–803 (2004).

To comment on this subject visit

http://blogs.nature.com/news/blog/2006/08/pros_and_cons_of_egg_donation.html

BOOKS & ARTS



Shock result? William Shockley (centre) shared the 1956 physics Nobel with John Bardeen (left) and Walter Brattain for their work on transistors.

The Moses of Silicon Valley

William Shockley's work led to the foundation of the US computer industry.

Broken Genius: The Rise and Fall of William Shockley, Creator of the Electronic Age

by Joel N. Shurkin

Macmillan Science: 2006. 400 pp. £19.99, \$27.95

Paul Grant

William Shockley was arguably the most enigmatic, provocative and controversial twentieth-century US physicist. His notoriety derived both from his perceived indirect role in the invention of the transistor, alongside his fellow 1956 Nobel laureates John Bardeen and Walter Brattain, and from his immersion in later life in the quagmire of the 'race intelligence' controversy. During his productive years as a physicist, Shockley excelled in collecting and exploiting ideas, both his own and those of his colleagues, but he failed dismally in exercising the leadership skills required to bring them to practical fruition. He has been called by some the 'Moses of Silicon Valley' — a fitting description in that he lured to the 'promised land' those who built the technical foundation for today's Information Age. But, like Moses, Shockley did not himself enter the land of milk and honey.

In the aptly titled biography *Broken Genius*, Joel Shurkin reveals Shockley to be a fascinating example of an aristotelian tragic hero whose flaw is readily discernible in the first act. In Shockley's case, this flaw was his overriding confidence in his own intelligence.

Shockley was an only child, born in 1910 in London to a gun-slinging mother and an MIT-educated father, a mining engineer whose search for gainful employment caused the family incessant relocation. As a child Shockley exhibited very disruptive behaviour at home, but was much better controlled in public (a glimpse of what was to come?). After his father's death in the 1920s, the family returned to the United States and settled in southern California. After high school, Shockley went to the California Institute of Technology and later MIT in the late 1920s and early 1930s during the 'golden age' of physics. He was heavily influenced by the epistemological debates of the time over determinism versus probability, which may well have fuelled his later attraction to statistical techniques.

After obtaining his PhD from MIT in 1936, Shockley was recruited by Mervin Kelly, director of Bell Telephone Laboratories. Then, as

now, it was unusual for a newly minted PhD to go to an industrial laboratory. It was perhaps an indication that Shockley wanted his life to have some practical impact. Besides, it was the middle of the Great Depression, and Shockley now had a wife and family to support.

Kelly saw that the electromechanical relays that pervaded central telephone switchgear at the time would not be able to handle the ever-increasing traffic load. He asked his Bell Labs team to find an alternative. Although the 1930s was the heyday of the vacuum-tube amplifier, Shockley thought it might be possible to leapfrog that technology and find some 'solid state' effect that could act as a switching device.

Shurkin's tale of the events, technical and behavioural, that led to the successful fabrication of the point contact and junction transistor by the late 1940s is riveting, even if you know the story already. Much has been said about who did what and how credit should have been properly apportioned. I once asked Bardeen this question, and he gave his usual reply: "The answer is in the preface to Shockley's book." There, in the opening pages of his 1950 textbook *Electrons and Holes in Semiconductors*,

Shockley attributes the invention of the transistor to Bardeen and Brattain, and to them alone. Bardeen had the key idea of minority carrier injection that made amplification possible, and Brattain had the skills to put the contacts close together. Yet without Shockley's participation and leadership, it is equally clear that the invention would not have occurred as soon as it did, and his later independent invention of the junction transistor was to emerge as the most practical embodiment of the device for the next decade.

Almost immediately after the discovery of the point-contact transistor, Shockley dissociated himself from many of his colleagues at Bell Labs, and eventually became disenchanted with the institution itself. Shurkin hints that this was the result of jealousy at not being fully involved in the final, crucial point-contact transistor experiments, and frustration at not progressing rapidly up the laboratory management chain. He had, in the words of his employees, an "unusual" management style.

For much of the early 1950s, Shockley was on leave from Bell Labs. His time was divided between teaching at Caltech and continuing to explore the statistical methods he introduced as a consultant to the military to help optimize naval and air-force tactical procedures in the Second World War. He adapted operations research techniques with the objective of maximizing damage to the enemy with the least expenditure in blood and money on his own side, a cost-benefit approach to conducting warfare.

Shockley's important wartime contributions have remained largely unknown, and Shurkin provides a rare focus on them. Unlike his better-publicized peers on the Manhattan Project and at MIT's Radiation Laboratory, Shockley was virtually on the battlefield and his recommendations had almost immediate effect. He was undoubtedly responsible for preventing many Allied casualties. He seemed to thrive in a leadership role within the command-and-control military culture. In some bizarre sense, he may have been cast in the mould of George Patton, Viscount Montgomery and Ulysses S. Grant — superb leaders of men in war, yet mostly dismal failures in times of peace.

After several years searching for an alternative career, Shockley finally left Bell Labs in 1956, returning to California to start the Shockley Semiconductor Laboratory, with financial backing from the industrialist Arnold Beckman. Once more, Shockley displayed an extraordinary gift for recruiting talent, along with total incompetence in managing them. Unlike Kelly, who gave his stable of thoroughbreds their head, Shockley attempted to micromanage, forcing his agenda on his new hires, whose ideas were in fact much better than his own. In 1957, Shockley's staff revolted, with eight — the 'traitorous eight' — resigning en masse. They were reunited by the inventor and financier Sherman Fairchild, and Fairchild Semiconductor was born. By 1961, Shockley

NEW IN PAPERBACK

Warped Passages: Unravelling the Universe's Hidden Dimensions

by Lisa Randall (Penguin, £8.99)

"The book's peg — the possible existence of extra dimensions in space — is easy enough to explain. But motivating the conjecture requires a grand tour of some of the toughest and most abstract topics in science." Paul Davies, *Nature* **435**, 1161-1162 (2005).

Madame Bovary's Ovaries: A Darwinian Look at Literature

by David P. Barash & Nanelle R. Barash (Delta/Bantam Dell, \$14)

"Where most such books rely on scholarly papers and monographs to ground the various points within a shared and robust scientific knowledge [of evolutionary biology], these authors use literature sources as references." Michel Raymond, *Nature* **435**, 28 (2005).

Oxford Dictionary of Scientific Quotations

by W.F. Bynum & Roy Porter (Oxford University Press, £10.99)

Semiconductor had folded. Seven years later, two of the eight, Robert Noyce and Gordon Moore, with financial help from the other six, founded Intel.

Shockley's consolation prize was a professorship at Stanford, which gave him time to pursue his fascination with the possible connection between heredity and intelligence. This

soon morphed into his much-sensationalized claim that African-Americans are statistically inferior in intelligence to those of European descent.

Had Shockley not been a Nobel laureate, his assertions would probably have been ignored. But a Nobel prize confers a 'bully pulpit', along with the perception, by people in general and the press in particular, of being an expert in everything. However, as the physicist and pundit Bob Park has observed, "A Nobel prize in physics is not an inoculation against silly behaviour," and there are plenty of examples to back that up.

With the discovery of DNA, physics met genetics, and the full impact of their engagement is just beginning to be glimpsed. If someday a gene sequence is found that determines intelligence, just as there are sequences for Tay-Sachs disease or thalassaemia, what shall we do about it? Will we have a future, as depicted in the film *Gattaca*, in which parents can selectively tailor the genetic structure of their children? If most choose intelligence as the dominant gene for their offspring, we might end up with a world full of Shockleys. Hopefully, this brave new world will have its share of those with the grace under pressure of Tiger Woods, the spine-tingling voice of Luciano Pavarotti, and the flashing fingers of rock guitarist Angus Young.

Variety is indeed the spice of life. ■
Paul Grant is an IBM research staff member emeritus and is currently visiting scholar in applied physics at Stanford University in California.

Making sense of autism

Autism, Brain, and Environment

by Richard Lathe

Jessica Kingsley: 2006. 288 pp.
£15.99, \$24.95

Understanding Autism: From Basic Neuroscience to Treatment

edited by Steven O. Moldin & John L. R. Rubenstein

CRC Press: 2006. 526 pp. £92, \$159.95

Francesca Happé

Are we witnessing an autism epidemic? The current prevalence estimates for autism, and for the wider range of autism spectrum disorders (ASDs), are around ten times greater than estimates from studies in the 1980s and 90s. About 0.6% of the population is thought to have an ASD, diagnosed on the basis of qualitative impairments in social interaction and communication, with restricted and repetitive interests and activities.

The claim that cases of ASD are on the increase is the first step in Richard Lathe's argument in *Autism, Brain, and Environment*. Based on the apparent increase, Lathe argues for an environmental explanation for what he

terms "new phase autism". He recognizes the overwhelming evidence that autism is among the most heritable of psychiatric disorders, but argues for a two-hit mechanism, with genetic susceptibility and environmental factors combining to produce an ASD. His book is a clearly and accessibly written account of his proposal that environmental poisons, including heavy metals, interact with genetic vulnerability to cause damage to the limbic brain system and to physiological systems, including the gut and the immune system, resulting in autism.

This is a story that many readers will find plausible, and which Lathe supports with some good synthesis of established autism research. It is perhaps not surprising that he also cites less solid, unpublished research to support the hypothesis, nor that the limitations of such research, such as the lack of appropriate control groups, are little discussed. But this is, overall, a scholarly book providing a possible explanation of autism. It will be of interest to parents as well as professionals.

Lathe's story stands in marked contrast to *Understanding Autism*, a volume edited by

Steven Moldin and John Rubenstein. This excellent collection of chapters on ASD, and on basic research relevant to ASD, provides few answers, but, I fear, better reflects the truth about our understanding of autism. A detailed and thorough examination of epidemiological studies by Eric Fombonne, for example, casts real doubt on the claim that the rate of ASD cases is increasing. Instead, he concludes that truly comparable time-trend data do not exist; until these are collected, changes in diagnostic criteria and practices, greater awareness and improved services may be sufficient to explain the increase in identified cases. And if there is no big rise in prevalence, there is no need to invoke a major environmental contribution.

Similarly, the idea that autism can be explained by damage to the limbic system (the

'social brain') is probably too simplistic. The chapter by Schumann, Bauman, Machado and Amaral reviews evidence from human and animal lesion studies. They suggest that the amygdala is not the seat of social processing, but instead plays a role in regulating fear behaviour, with knock-on effects exacerbating social deficits in autism. Indeed, many examples in the book illustrate how much more complex the story of autism is likely to be, including epigenetic regulation (not involving differences in the nucleotide sequence) through DNA methylation or chromatin remodelling.

Should we be despondent in the face of such complexity and the paucity of answers regarding the causes of autism? I think the book by Moldin and Rubenstein gives as much reason for hope as for despair. The chapters reflect the

diverse tools being called into the service of understanding autism, including neuroimaging, animal models, even the use of neurotropic viruses, which attack the nervous system, to trace chains of synaptically linked neurons.

Parents, teachers, clinicians and researchers are all searching for explanations, for a story that will help us make sense of autism and show us how to help. Lathe's book tells a fascinating tale, but Moldin and Rubenstein's may be closer to the truth. After all, the one thing that all those who study autism agree on is that no one really understands autism. ■

Francesca Happé is a reader in cognitive neuroscience at the MRC Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, King's College London, London SE5 8AF, UK.

Subterranean storage blues

Uncertainty Underground: Yucca Mountain and the Nation's High-Level Nuclear Waste
edited by Allison M. Macfarlane & Rodney C. Ewing
MIT Press: 2006. 416 pp. \$72, £46.95 (hbk); \$29, £18.95 (pbk)

Gordon MacKerron

Earth scientists generally have strong views about the deep disposal of radioactive waste. They all agree that deep disposal is the safest option, and that several suitable geologies exist, but it is difficult to gain consensus about the safety of any specific site. *Uncertainty Underground*, edited by Allison Macfarlane and Rodney Ewing, displays these viewpoints clearly. The individual chapters are not only well-written but also broadly comprehensible, even to a social scientist such as me. All the authors — who belong mostly to the Earth-sciences community and work variously in academia, industry, government and consultancy — are in favour of deep disposal. But they hold very different views on the suitability of the Yucca Mountain site in Nevada, which has been the chosen resting place for US high-level waste and spent nuclear fuel since 1987.

The editors have had an unusually strong influence on this collection. They selected authors and subjects to illustrate their main theme — that too much attention has been paid to the 'near-field' (broadly, the engineering of the repository and its contents) and not enough to the less-controllable issues of the geology and hydrogeology of Yucca Mountain and its surroundings.

The Yucca Mountain site is, in international terms, unique as a potential repository site because it is 'dry', being well below the surface of the mountain but, critically, well above the water table. Some authors argue that this 'advantage' is potentially problematic: hydrogeologists have concentrated on 'wet' sites, and have done little work on the hydrology of dry sites. Potentially major climatic, volcanic,

uncertain (transport in the saturated zone).

Another section covers waste packages and waste forms, for which the uncertainties seem considerably smaller (as the editors suggest) than for the geology and hydrogeology. However, a long-running debate rumbles on between advocates of glass versus ceramics as the best medium in which to contain high-level wastes.

For social scientists, the sections on policy and uncertainty are especially interesting. The debates about the standard of 'proof' of safety over timescales of up to a million years are well and clearly presented. Particularly good are the sections on the distinction between risk (the probability of something going wrong, from which decision-making at least has a starting point) and the inevitable uncertainties over very long periods into the future, for which precise risk levels are unknown and probably unknowable. This raises the issues of ethics and the way the rights of distant generations are balanced against those of the near future, although these are not covered here.

Finally, on policy, it is clear that the choice of Yucca Mountain, especially in the legislation of 1987, is severely clouded by a lack of political legitimacy that may yet prove highly problematic.

This book shows clearly that a site such as Yucca Mountain can probably be made safe enough if the geology is sufficiently well understood and married to the right engineering. But the question of whether it will be perceived by the public as a good site depends on a political process that has, at best, been poorly handled in the United States. ■

Gordon MacKerron is chair of the UK Committee on Radioactive Waste Management, and director of the Sussex Energy Group, SPRU, University of Sussex, Brighton BN1 9RH, UK.



Deep divisions: the disposal of nuclear waste at Yucca Mountain has split opinion.

seismic and other changes over periods of up to a million years in the future could in principle lead to substantially different hydrological regimes, with evident long-term safety implications.

Many of the papers in this volume therefore review specific issues of hydrology and thermohydrology at Yucca, such as the risk of hot upwelling water from the water table flooding the site in the future, radionuclide transport in the unsaturated zone, and contaminant transport in the saturated zone. Varying degrees of reassurance emerge from this hydrology-based set of papers, from virtually complete (with regard to hot upwelling) to much more

NEWS & VIEWS



A. JAVELLANA

Figure 1 | Inundated rice fields in the Philippines. The submergence of rice plants during flooding has severe consequences for rice production.

PLANT BREEDING

Rice in deep water

Takuji Sasaki

Many otherwise productive cultivars of rice suffer badly if immersed in water for long. The identification of a gene variant that confers tolerance to this threat has practical potential.

Like all crops, rice plants of course require water to grow. But you can have too much of a good thing: when excessive water results in prolonged submergence, the effects on rice production can be devastating. Hence the significance of Xu and colleagues' investigation of the genetics of submergence tolerance, described on page 705 of this issue¹.

Rice is an indispensable staple food, especially in the large areas of Asia, Latin America and Africa that are characterized by a semi-tropical climate with alternating rainy and dry seasons. Since the beginning of agriculture, rice plants have been adapted for a wide range of environmental conditions through continuous breeding and selection. Worldwide, more than 120,000 rice varieties have been bred to satisfy demands ranging from high yield potential, disease resistance, tolerance to stresses, good eating quality and increased nutritional value. Two of the best-known success stories are the introduction of semi-dwarfism in the late 1960s by the International Rice Research Institute in the Philippines, which propelled the Green Revolution², and the development of hybrid rice technology in the mid-1970s in Hunan, China³. Together,

these developments have had astounding results: rice production has doubled over the past 40 years.

Nonetheless, demand for rice exceeds production, a situation that will get worse given the world's increasing population and the decline in extent of arable land. Production in regions where rice cultivation is subjected to stresses such as the seasonal flooding that occurs during the monsoon season, particularly the lowlands of south, southeast and east Asia (Fig. 1), is especially unpredictable. The rivers of these regions further contribute to annual flooding of low-lying agricultural lands. Floods during and immediately following the rainy season can submerge young rice plants for several days, resulting in wilting and often death.

This is where Xu and colleagues¹ come in, for they have characterized a major rice gene that confers tolerance to prolonged submergence. They used a conventional genetic mapping approach to identify a stretch of DNA — a 'quantitative trait locus' — associated with immersion tolerance in a particular rice cultivar that has that trait. This locus, *Submergence 1* (*Sub1*), occurs on chromosome 9. Xu *et al.*

found that three genes of the ethylene-response-factor (ERF) family are sequentially arrayed in this locus. A variant of one of them, *Sub1A-1*, was found only in submergence-tolerant rice and could induce intolerant varieties to survive prolonged periods of immersion.

The authors further discovered that a 'single nucleotide polymorphism' distinguishes *Sub1A-1* from the intolerant (*Sub1A-2*) variant of the gene; this difference results in the replacement of the proline amino acid with serine in part of the ERF protein known as the MAPK site. The ERF protein is a plant-specific gene-regulatory factor, and the phosphorylation of serine in the MAPK site helps to control the transcription of target genes such as alcohol dehydrogenase (*Adh1*) that are normally activated in anaerobic conditions (as apply when a plant is submerged). This line of reasoning is supported by Xu and colleagues' finding that overexpression of *Sub1A-1* in intolerant cultivars conferred tolerance to submergence as well as enhancing *Adh1* activity.

The ERF family includes genes associated with disease-related responses; for example, the production of signalling molecules such

as ethylene, jasmonic acid and salicylic acid. These genes are similar to the 'dehydration-responsive element binding' genes, which are involved in countering abiotic stresses such as dehydration or cold⁴. Both types of gene encode a conserved domain called AP2, a unique plant gene-regulatory factor⁵. The genomic characterization of submergence tolerance based on *Sub1A* may therefore help to clarify the mechanisms involved in tolerance to other abiotic stresses.

The particular impact of this study¹ lies in Xu and colleagues' accurate and effective introduction (introgression) of *Sub1A-1* into local rice varieties subject to seasonal flooding, as well as into elite and semi-dwarf varieties normally grown in intensively irrigated areas

that are also vulnerable to flooding. Successful introgression of *Sub1A-1* was achieved by a technique called marker-assisted selection. This is a crucial point: the technique enables plant breeders to overcome a phenomenon known as linkage drag, in which deleterious genes may be incorporated into the target cultivar along with the beneficial genes.

The success in characterizing the *Sub1A* gene demonstrates the value of having a high-quality reference sequence from a single plant cultivar for accurate detection of genetic variation. Knowledge of the sequence of specific genes and their associated variants will enable researchers to tap into the natural genetic variation in a wide collection of rice germ lines. The 'gold standard' sequence data⁶ generated

by the International Rice Genome Sequencing Project will really begin to prove its value as breeders and geneticists hunt down the elusive genes that protect rice plants against all types of stress.

Takuji Sasaki is at the National Institute of Agrobiological Sciences, 1-2, Kannondai 2-chome, Tsukuba, Ibaraki 305-8602, Japan.
e-mail: tsasaki@nias.affrc.go.jp

1. Xu, K. *et al.* *Nature* **442**, 705–708 (2006).
2. Peng, S. *et al.* *Crop Sci.* **39**, 1552–1559 (1999).
3. Yuan, L. P. in *Hybrid Rice Technology: New Developments and Future Prospects* (ed. Virmani, S. S.) 1–6 (Int. Rice Res. Inst., Los Baños, Philippines, 1994).
4. Kasuga, M. *et al.* *Nature Biotechnol.* **19**, 287–291 (1999).
5. Riechman, J. L. *et al.* *Science* **290**, 2105–2110 (2000).
6. International Rice Genome Sequencing Project *Nature* **436**, 793–800 (2005).

COSMOLOGY

Where all the lithium went

Corinne Charbonnel

For some years, astronomers have been trying to track down all the lithium predicted by standard cosmological models. Spectroscopic dissection of globular clusters reveals that the answer might lie in the stars.

Cosmologists have a problem. They have highly precise models of how light elements were synthesized in the immediate aftermath of the Big Bang. But when they compare their estimates of the abundance of lithium with the amounts that astronomers observe in the oldest stars of our Galaxy, they find that their predictions are far too high. So who's bluffing?

On page 657 of this issue, Korn *et al.*¹ enter as arbiters. They report the detection of diffusion signatures in the abundances of the lithium isotope ⁷Li and iron in stars of a globular cluster — a tightly bound collection of up to half a million old stars — known as NGC 6397 (Fig. 1). These challenging observations were performed with the European Southern Observatory's Very Large Telescope in the Chilean Atacama Desert. They seem to give a clear answer to the lithium problem: the old stars have simply destroyed part of their pristine ⁷Li. That insight could restore confidence in the standard picture of Big Bang nucleosynthesis, and give fresh impetus to researchers working to increase the sophistication of stellar models.

Together with hydrogen, deuterium and the two helium isotopes ³He and ⁴He, ⁷Li is one of the few nuclides that were synthesized in significant quantities when, briefly, the newborn Universe functioned as a cosmic nuclear-fusion reactor. The relic of this primordial lithium was discovered 25 years ago in the atmosphere of the oldest stars of our Galaxy². Remarkably, those stars in the Galactic halo were found to have a constant ⁷Li abundance that was independent of their mass and metal-

licity (the proportion of elements other than hydrogen and helium in the star's make-up). That uniformity was interpreted as proving the cosmological origin of ⁷Li.

Then, a few years ago, the analysis of unevenness found in the cosmic microwave background³ — the cooled primal glow of the Big Bang — constrained the predictions of nucleosynthesis in the early Universe with unprecedented precision⁴. But the analysis

yielded a theoretical abundance of ⁷Li about three times higher than that observed in halo stars. Since then, astrophysicists and cosmologists have raced to resolve this difference.

One logical explanation for the dilemma is that the ⁷Li found in halo stars is not pristine. The isotope is fragile, and burns in stellar interiors where the temperature exceeds 2.5 million kelvin. It could, therefore, have undergone *in situ* depletion through hydrodynamic processes that modify the chemical composition of stars throughout their life.

This simple idea was advocated long ago⁵ based on observations of the abundance of lithium in the Sun and in younger solar-type stars in our Galaxy, which are known to slowly deplete their initial lithium stock. Given that halo stars are older than solar-type stars, one might posit such depletion as the explanation for their low ⁷Li content. This interpretation was however dismissed, because *ad hoc*



Figure 1 | Ancient relic. The globular cluster NGC 6397 contains around 400,000 stars and is found some 7,200 light years away in the southern constellation Ara. Together with halo stars, globular-cluster stars are among the oldest objects in the Milky Way, and hold the nuclear relics of the Big Bang. Evidence for complex internal hydrodynamic processes that reduce these stars' lithium abundance could explain the cosmological lithium discrepancy, as Korn and colleagues¹ reveal.

ESO

assumptions had to be made in order to explain why the ^7Li content of halo stars of varying mass and metallicity is the same.

A star is essentially a gaseous mixture of hydrogen and heavier elements, and the stellar gas as a whole cannot be in equilibrium until each component reaches its own equilibrium distribution. The equilibration occurs through microscopic processes, collectively known as atomic diffusion, that encompass gravitational, thermal and radiative effects. These diffusive effects cause ^7Li and metals such as iron to sink from the atmosphere of a halo star and hide below the star's surface.

However, this effect increases with stellar mass, in contradiction to the observation in halo stars. That seemed to suggest that such mechanisms are inefficient in these stars. In fact, this difficulty was probably telling us simply that our understanding of stellar hydrodynamics was still rather poor.

Specifically, the uniform reduced lithium abundance in halo stars can be interpreted as the signature of macroscopic processes resulting, for instance, from the rotation of the star and internal gravity waves. Such processes would minimize the effects of atomic diffusion, but would direct ^7Li from the surface of the star towards its hotter interior, where it would burn up. Evidence for these complex mechanisms has already been found in the Sun through helioseismology⁶ — the study of wave oscillations in the Sun's interior — and from lithium data in halo⁷ and younger⁸ stars.

Korn *et al.*¹ looked for signatures of such lithium depletion processes in the stars of the NGC 6397 cluster. The stars of a globular cluster are old, and are very similar to halo stars. In particular, they are also thought to have been born with the cosmological lithium abundance. The stars are coeval (of the same age), and their mass and evolutionary stage can be determined unambiguously. This last point is essential for determining an evolutionary effect on the stellar surface abundances. But the stars' faintness, together with the high analytical accuracy required to detect star-to-star abundance variations, demands a large amount of observing time — even on telescopes with 8-metre apertures such as those of the Very Large Telescope. Thus, Korn and colleagues' spectroscopic dissection could be performed for only a few stars.

Nevertheless, the results are unambiguous and consistent with previous observational evidence from lithium data in halo⁷ stars. They show that, as the observed globular-cluster stars age and cool, the proportion of lithium in their atmosphere first slightly increases and then drops off sharply. Over the same temperature range, the proportion of iron increases steadily, albeit more slowly. Both trends can be interpreted as clear signatures of transport processes within the star that compete with atomic diffusion and lead to the destruction of lithium. More precisely, the data can be nicely reproduced by stellar models⁹ that include the

effects of atomic diffusion in the presence of weak turbulence — although the origin of that turbulence remains uncertain. Extrapolating backwards in time, the initial lithium content of globular-cluster and halo stars can be inferred, and the value agrees very tidily with the predictions from Big Bang nucleosynthesis.

So although, thanks to Korn *et al.*¹, the cosmological lithium crisis is probably over, the race for answers has merely changed its objective. Observers have provided us theorists with superb constraints for our stellar models. We must now unambiguously determine the nature of the turbulent process that weakens atomic diffusion and leads to the constant lithium abundance measured in most halo stars. Sophisticated models that take into account hydrodynamic processes induced by stellar rotation and internal gravity waves have been used to explain the lithium patterns observed in the Sun and in young solar-type stars¹⁰. We must urgently check whether they

can also explain the lithium content of the galactic halo stars.

Corinne Charbonnel is at the Observatoire de Genève, 51 chemin des Maillettes, 1290 Sauverny, Switzerland, and the Observatoire Midi-Pyrénées (CNRS UMR 5572), Toulouse, France.
e-mail: Corinne.Charbonnel@obs.unige.ch

1. Korn, A. J. *et al.* *Nature* **442**, 657–659 (2006).
2. Spite, M. & Spite, F. *Nature* **297**, 483–485 (1982).
3. Spiegel, D. N. *et al.* *Astrophys. J. Suppl. Ser.* **148**, 175–194 (2003).
4. Coc, A. *et al.* *Astrophys. J.* **600**, 544–552 (2004).
5. Vauclair, S. *Astrophys. J.* **335**, 971–975 (1988).
6. Guzik, J. A. & Cox, A. N. *Astrophys. J.* **386**, 729–733 (1992).
7. Charbonnel, C. & Primas, F. *Astron. Astrophys.* **442**, 961–992 (2005).
8. Charbonneau, P. & Michaud, G. *Astrophys. J.* **334**, 746–760 (1988).
9. Richard, O., Michaud, G. & Richer, J. *Astrophys. J.* **619**, 538–548 (2005).
10. Charbonnel, C. & Talon, S. *Science* **339**, 2189–2191 (2005).

NEUROSCIENCE

An extra dimension to olfaction

John Ngai

The sense of smell is triggered by receptors in the olfactory epithelium that lines the nose. In mice at least, that lining is also responsible for receiving chemosensory cues involved in mating and other social behaviours.

In 1991, Linda Buck and Richard Axel reported the seminal discovery of the gene family that encodes odorant receptors in vertebrates¹. A paper by Buck and Stephen Liberles² on page 645 of this issue describes a second class of chemosensory receptor expressed by olfactory sensory neurons*. Adding another dimension to a rapidly progressing field, this unexpected discovery provides new insights into the molecular and cellular mechanisms used to detect olfactory cues.

Since 1991, we have learned a great deal about how the vertebrate olfactory system receives and processes sensory information from myriad chemical cues. The initial sensors are receptors, known as G-protein-coupled receptors (GPCRs), which number more than 1,000 in some mammalian species and are instrumental in the detection of tens, if not hundreds, of thousands of volatile odorants. In parallel, the vomeronasal organ — an anatomical specialization of the nose in terrestrial vertebrates that is distinct from the main olfactory epithelium — senses non-volatile chemical stimuli, including pheromones³. Detection of chemosensory stimuli by the vomeronasal organ is thought to be

mediated by two other families of GPCRs that function as pheromone receptors⁴.

Liberles and Buck set out to answer a simple yet important question: can the known family of olfactory receptors account for all the effects and behaviours that are mediated by olfaction? They used a large-scale screening approach, known as 'quantitative reverse transcription-polymerase chain reaction' (Q-PCR), to survey mouse olfactory neurons for the expression of other GPCRs identified from the mouse genome. They found that members of a particular group of receptors, the trace amine-associated receptor (TAAR) family⁵, are expressed in these neurons. TAARs were originally identified in a search for amine receptors in the brain⁶ and show sequence similarities to receptors for the neurotransmitters serotonin and dopamine. Members of the TAAR family are activated by the trace amines found in the central nervous system (β -phenylethylamine, tyramine, tryptamine and octopamine), and individual TAARs are expressed in small numbers of cells in the brain. Trace amines have long been suspected to be involved in psychiatric disorders, so TAARs have been postulated to play a role in depression and schizophrenia⁵.

Liberles and Buck² now provide a different perspective on TAAR function. Using Q-PCR, they found that eight out of nine

*This article and the paper concerned² were published online on 30 July 2006.

mouse TAAR subtypes are expressed in the olfactory epithelium, whereas none was detected in the brain. Although TAARs might be expressed in too few brain cells to be detected by this assay, at face value these results suggest that a major function of TAARs is the detection of olfactory cues by the nose. Remarkably, the authors also show that messenger RNAs for individual TAARs are expressed in a pattern in the olfactory epithelium that is reminiscent of that displayed by the canonical odorant receptors: individual receptors are sparsely expressed in discrete subdomains of the epithelium, and are not co-expressed with other TAARs. They are also probably not expressed with the odorant receptors. So, based on their expression patterns alone, TAARs seem to fit the bill as chemosensory receptors.

What chemical cues do TAARs receive? To answer this question, Liberles and Buck screened more than 300 odorous compounds (including almost 100 amines) for their ability to activate tissue culture cells engineered to express TAARs. They identified activating compounds for five of the receptors tested; not surprisingly, all of them were amines, including those shown previously to activate TAARs. Individual TAARs are specific for different amine structures; three of them were activated by volatile amines found in urine (a source of pheromonal cues of a variety of chemical compositions), some of which have been implicated in regulating reproductive behaviour. Interestingly, the authors found that urine from sexually mature male mice — but not from females or sexually immature males — could stimulate mTAAR5, a receptor activated by trimethylamine. Taken together, these results suggest that, in mice, TAARs may mediate behavioural and physiological responses to amine-based social cues present in urine. But it remains to be seen whether TAARs are indeed required for any stereotyped reproductive or other social behaviours.

Recent studies have highlighted how both the main olfactory epithelium and the vomeronasal organ operate in receiving pheromonal cues^{7–10}. Liberles and Buck's results underscore the importance of the main olfactory system in this process and provide the molecular tools to tackle questions about the pathways subserving olfactory-mediated behaviours.

In the processing of information from odorant receptors, sensory neurons expressing the same receptors converge on the same sites (glomeruli) of a relay station called the olfactory bulb, forming a spatial map of sensory information. From there, mitral cells send information on to another part of the brain, the olfactory cortex. Do sensory neurons expressing individual TAARs likewise converge on common glomeruli in the olfactory bulb? And do the mitral cells that innervate these glomeruli in turn project to

regions of the olfactory cortex that are distinct from those serving other olfactory pathways? The expression of individual TAAR genes in small subsets of neurons also raises questions about the regulation of these genes — such as whether the mechanisms that are used to select a single gene variant to encode an odorant receptor in each sensory neuron¹¹ are also used to regulate TAAR gene expression.

More generally, the genes that encode TAARs are found in diverse vertebrate genomes, including those of humans and certain fish¹². Evolutionary biologists, as well as physiologists and those studying animal behaviour, will be curious to find out how TAARs evolved for communication in other vertebrate species. ■

John Ngai is in the Department of Molecular and Cell Biology, and the Helen Wills Neuroscience

Institute, University of California, Berkeley, California 94720, USA.

e-mail: jngai@socrates.berkeley.edu

1. Buck, L. & Axel, R. *Cell* **65**, 175–187 (1991).
2. Liberles, S. D. & Buck, L. B. *Nature* **442**, 645–650 (2006).
3. Dulac, C. & Torello, A. T. *Nature Rev. Neurosci.* **4**, 551–562 (2003).
4. Mombaerts, P. *Nature Rev. Neurosci.* **5**, 263–278 (2004).
5. Lindemann, L. & Hoener, M. C. *Trends Pharmacol. Sci.* **26**, 274–281 (2005).
6. Borowsky, B. et al. *Proc. Natl Acad. Sci. USA* **98**, 8966–8971 (2001).
7. Lin, D. Y., Zhang, S. Z., Block, E. & Katz, L. C. *Nature* **434**, 470–477 (2005).
8. Mandiyan, V. S., Coats, J. K. & Shah, N. M. *Nature Neurosci.* **8**, 1660–1662 (2005).
9. Spehr, M. et al. *J. Neurosci.* **26**, 1961–1970 (2006).
10. Wang, Z. et al. *J. Neurosci.* **26**, 7375–7379 (2006).
11. Chess, A., Simon, I., Cedar, H. & Axel, R. *Cell* **78**, 823–834 (1994).
12. Gloriam, D. E. et al. *Mol. Phylogenet. Evol.* **35**, 470–482 (2005).

MATERIALS SCIENCE

Organosilica the conciliator

Mietek Jaroniec

Acidic and basic molecules are antagonistic, and keeping them in their place is no easy job — unless, it seems, one unites them under the tutelage of ordered, nanoporous materials known as organosilicas.

Silica (silicon dioxide, SiO₂) is the basic component of sand and quartz, and the most abundant compound in Earth's crust. It can also be synthesized as an amorphous solid with nanoscale pores and an accessible internal surface of up to 1,000 square metres per gram (ref. 1). Combining such a carefully designed porous silica framework with various organic

molecules results in hybrid 'organosilica' materials that are capable of some impressive chemical feats. For instance, Johan Alauzun and colleagues² describe in the *Journal of the American Chemical Society* how they fabricated a mesoporous organosilica that allows two mutually antagonistic molecular groups — one acidic, one basic — to coexist peacefully.

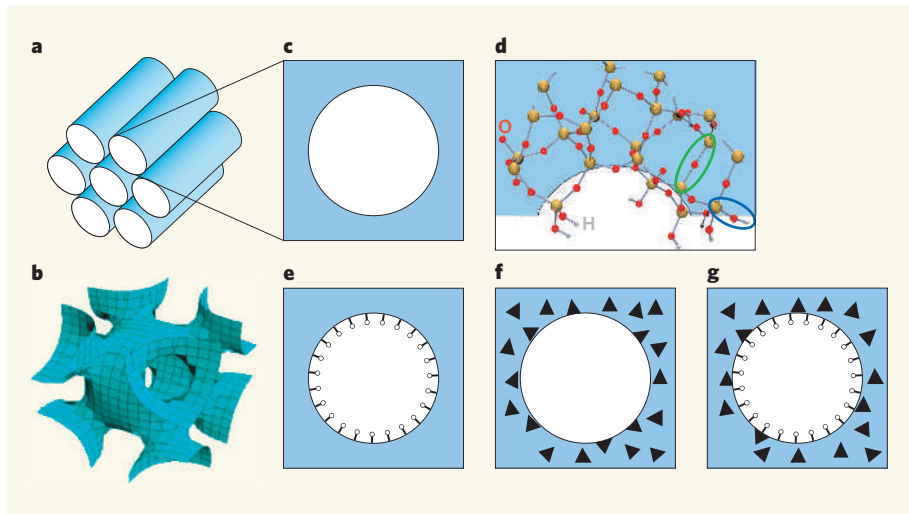


Figure 1 | Silica channels and cages. **a, b**, Mesoporous silica structures can be fabricated in channel-like (**a**) and cage-like (**b**) conformations. **c, d**, A cross-section through purely siliceous pore walls (**c**) reveals a structure (**d**) consisting of interconnected silicon (Si; yellow), oxygen (O; red) and hydrogen (H; grey) atoms. A siloxane (green ellipse) and a silanol (blue ellipse) group are highlighted. **e–g**, Mesoporous silica structures can accommodate organic groups in several ways: **e**, pores decorated with organic surface groups; **f**, an outer framework with organic 'bridging' groups; **g**, bifunctionality with both surface and bridging groups.

Silica nanostructures can be easily prepared by allowing a silicate precursor material to self-assemble on an organic template gel. When the template is removed or combusted, channel-like or cage-like pores are left behind, and other molecular groups can be attached or immobilized here^{1,3} (Fig. 1a, b). In both cases, the pore walls of silica nanostructures consist of randomly connected SiO₄ tetrahedra. In each of these, a silicon atom forms four bonds, one to each of four oxygen atoms. As an oxygen atom has two available bonds, it connects to two silicon atoms. Thus, in the bulk of the framework, each silicon atom is bound to four neighbours through oxygen atoms (Fig. 1d). The exception is on the surface of the framework, where the oxygen atoms use one of their two bonds to link up with a hydrogen atom. Chemically speaking, the characteristic features of silica materials are the presence of siloxane groups (denoted $\equiv\text{Si}-\text{O}-\text{Si}\equiv$) in the bulk, and both silanol ($\equiv\text{Si}-\text{OH}$) and siloxane groups on the surface.

These reactive silanol groups are crucial, as they permit modification of the silica surface if hydrogen is replaced by an organic group, denoted 'R'. But this surface modification, known as post-synthesis grafting⁴, is not the only way to decorate the pore walls of silica nanostructures with surface groups. 'One-pot synthesis'⁴ can also be used, which comprises stages of self-assembly and co-condensation, in the presence of an organic template, of the silicon-containing precursors Si(OEt)₄ and RSi(OEt)₃. The ethoxy group (EtO) in these precursors leaves the precursor during reaction with water (hydrolysis), permitting the condensation and formation of the framework.

The principal advantage of this self-assembly approach is that, by using slightly more complex precursors known as silsesquioxanes, for example (EtO)₃Si-R-Si(OEt)₃, organic groups can be introduced into the silica framework during templating synthesis, rather than being cobbled on to its surface after the event. So-called periodic mesoporous organosilicas (PMOs) are the result⁵⁻⁷, in which organic bridging groups ($\equiv\text{Si}-\text{R}-\text{Si}\equiv$) replace some siloxane bonds. (The term 'mesoporous' refers to nanostructures with pore sizes in the range 2–50 nm.) Thus, self-assembly of simple or mixed precursors presents almost unlimited opportunities for incorporating both surface and bridging groups into silica nanostructures (Fig. 1e–g).

There is a problem. Although the self-assembly approach seems quite general, it is unsuited for the simultaneous incorporation of two or more groups with antagonistic properties, such as acidic and basic groups, because the precursors of these groups would simply react during organosilica synthesis.

But Alauzun and colleagues² demonstrate a deft way of transforming the framework and surface groups of a specially designed bifunctional PMO into acidic and basic groups,

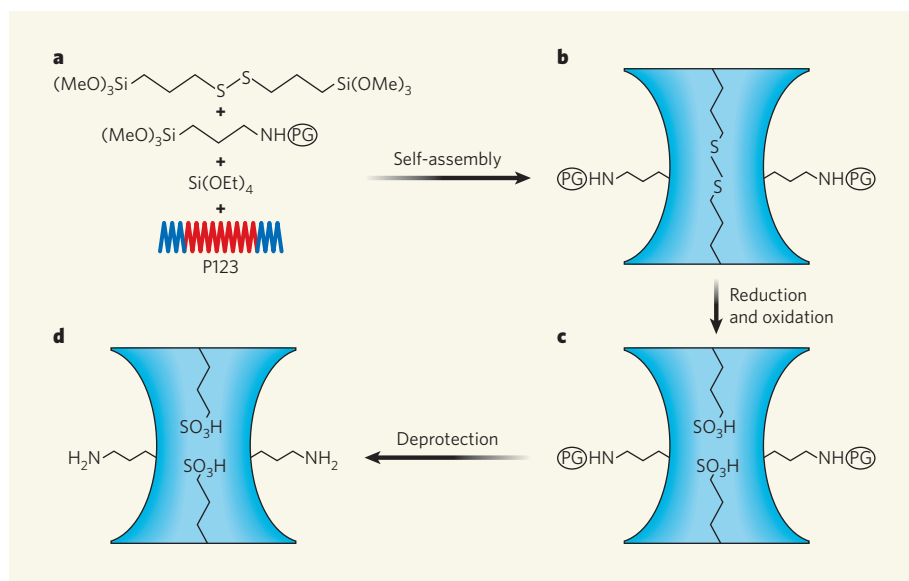


Figure 2 | Acidic framework, basic surface. **a**, Alauzun and colleagues' strategy² for the preparation of mesoporous organosilica materials with an acidic framework and basic pores first involves the self-assembly of two organosilanes containing, on the one hand, a disulphide ($-\text{S}-\text{S}-$) bridge and, on the other, a protected aminopropyl group, $\text{NH}(\text{PG})$. (OMe denotes a hydrolysable methoxy group, and OEt denotes an ethoxy group.) The self-assembly occurs in the presence of a triblock copolymer template (P123, containing two hydrophilic and one hydrophobic segments) under slightly acidic conditions. Protection of the basic aminopropyl group was achieved by adding *t*-butoxycarbonyl to this group to block its reactivity during organosilica preparation; at the final synthesis stage, the protecting group was removed to restore the basic functionality (deprotection). **b**, The resulting mesoporous material has disulphide bridges in the framework and protected aminopropyl groups on the surface. **c, d**, This material was subjected to a series of chemical reactions: **c**, reduction of the disulphide bridge to intermediate thiol groups followed by their oxidation to acidic sulphonhic groups (SO_3H) and **d**, deprotection of the protected aminopropyl to form a basic surface group.

respectively. The path to their results was paved with experimental studies on the synthesis of bifunctional PMOs (organosilicas containing both surface and bridging groups)⁸⁻¹¹, on chemical transformations of PMOs^{7,10}, and on the chemical reactivity of sol-gel materials with entrapped antagonist reagents¹².

First, the authors synthesized their bifunctional PMO by self-assembly of a disulphide-bridged organosilane (the acidic precursor), a protected aminopropyl group (the basic precursor) and tetraethoxysilane (used to consolidate the framework) in the presence of a block copolymer template (Fig. 2). They created acidic sulphonhic groups ($-\text{SO}_3\text{H}$) in the framework by converting disulphide bridges ($-\text{S}-\text{S}-$) to thiol ($-\text{SH}$) groups, initially under reducing conditions, and then oxidized these groups with hydrogen peroxide and dilute sulphuric acid. Finally, the protecting group, added to ensure that the functionality of the aminopropyl group is not altered during the chemical treatment of the bifunctional PMO, was removed by thermal treatment in vacuum conditions to obtain basic $-\text{NH}_2$ groups on the pore walls.

The authors confirmed the presence of an acidic framework by conducting a titration reaction of the PMO with sodium hydroxide, the archetypal aqueous base, and then showed that basic pores must also be present by titrating with aqueous hydrochloric acid. The

accessibility of the pores was tested in a condensation reaction that involved adding acrylamide ($\text{C}_3\text{H}_5\text{NO}$) to the basic aminopropyl groups. Elemental analysis of the resulting product showed that about 80% of basic groups had undergone the desired addition reaction.

Alauzun and colleagues' results² show that the self-assembly synthesis of mesoporous organosilicas with framework and surface groups, followed by appropriate chemical transformations, is a promising road towards obtaining nanostructures with two opposing properties. But the synthesis of materials with new antagonist groups is no easy task, as it requires the design and preparation of organosilica nanostructures that can be exposed to the appropriate chemical reactions without structural deterioration.

The rewards are nevertheless potentially great: Alauzun and colleagues' materials would, for instance, be ideally suited for the acid–base bifunctional catalysis that could be relevant to many industrial chemical reactions¹³. One could imagine chemical reactions in which a simultaneous ('concerted') interaction of the substrate molecule with both basic and acidic sites of the catalyst causes a shift of an electron pair, and a concomitant chemical transformation. Another possibility is that adjacent acidic and basic sites might activate two molecules of different substrates, and so kick-start their reaction. Such prospects

should stimulate others to attempt the conciliation of antagonists.

Mietek Jaroniec is in the Department of Chemistry, Kent State University, Kent, Ohio 44242, USA.
e-mail: jaroniec@kent.edu

1. Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. *Nature* **359**, 710–712 (1992).
2. Alauzun, J., Mehdi, A., Reyé, C. & Corriu, R. J. P. *J. Am. Chem. Soc.* **128**, 8718–8719 (2006).
3. Zhao, D. Y. *et al. Science* **279**, 548–552 (1998).
4. Stein, A., Melde, B. J. & Schrodin, R. C. *Adv. Mater.* **12**, 1403–1419 (2000).
5. Inagaki, S., Guan, S., Fukushima, Y., Ohsuna, T. & Terasaki, O. *J. Am. Chem. Soc.* **121**, 9611–9614 (1999).
6. Melde, B. J., Holland, B. T., Blanford, C. F. & Stein, A. *Chem. Mater.* **11**, 3302–3308 (1999).
7. Asefa, T., MacLachlan, M. J., Coombs, N. & Ozin, G. A. *Nature* **402**, 867–871 (1999).
8. Asefa, T. *et al. J. Am. Chem. Soc.* **123**, 8520–8530 (2001).
9. Burleigh, M. C., Markowitz, M. A., Spector, M. S. & Gaber, B. P. *Chem. Mater.* **13**, 4760–4766 (2001).
10. Yang, Q. H., Kapoor, M. P. & Inagaki, S. *J. Am. Chem. Soc.* **124**, 9694–9695 (2002).
11. Olkhoviyk, O., Pikus, S. & Jaroniec, M. *J. Mater. Chem.* **15**, 1517–1519 (2005).
12. Gelman, F., Blum, J. & Avnir, D. *Angew. Chem. Int. Edn* **40**, 3647–3649 (2001).
13. Setoyama, T. *Catal. Today* **116**, 250–262 (2006).

SOLAR SYSTEM

Sifting through the debris

Asantha Cooray

A quadrillion previously unnoticed small bodies beyond Neptune have been spotted as they dimmed X-rays from a distant source. Models of the dynamics of debris in the Solar System's suburbs must now be reworked.

Until the time of Galileo, the Solar System was believed to consist of the Sun, the Moon and the six inner planets — including Earth — as far out as Saturn. Since then, we have added other gas giant planets (Neptune and Uranus), moons of other planets, the asteroid belt and, in 1930, Pluto. On page 660 of this issue, Chang and colleagues¹ present observational estimates of the make-up of the Solar System's most distant components — the trans-neptunian objects. Their results call for a rethink of models of how these bodies formed.

It was in 1943 that Kenneth Edgeworth suggested the outer Solar System might contain a

large number of tiny 'planetesimal' objects². These objects are the residual debris from the planetary formation process and were swept off to the edge of the Solar System by the gravity of the giant planets. Since 1992, searches for slow-moving faint objects have revealed more than 50,000 planetoids, which are generally hundreds of kilometres across, orbiting beyond Neptune³. But despite theoretical models suggesting their existence, there had been no conclusive evidence for the presence of smaller bodies.

Chang *et al.*¹ change that. Their observations indicate that the trans-neptunian debris field in

fact contains many more small planetesimals of between 10 and 100 metres across than theorists had predicted. The authors infer this from random, rapid dips in the intensity of Scorpius X-1, the brightest source of X-ray light in the sky, which was monitored by NASA's Rossi X-ray Timing Explorer over a long period. Such dips can be caused by the passage of a dark object in the Solar System across the background source as viewed from Earth, a process known as occultation. An object 100 metres across at the edge of the Solar System could cause an occultation of a few milliseconds' duration. Such a small object so far away cannot be directly viewed with even the largest telescopes, as the amount of sunlight that it reflects back to Earth is minuscule.

The use of stellar occultations to reveal unknown facets of our Solar System is not new. Uranus's rings were discovered in this way, as was a host of details about the atmospheres of other planets and their satellites⁴. The idea of using occultations to search for small bodies in the outer Solar System has been around since 1976 (ref. 5). But in practice, most detectors that are used for astronomical observations at optical wavelengths cannot record changes at the sub-second time intervals required for this particular hunt.

Chang and colleagues' inspiration was to realize that detectors counting X-ray photons record light at shorter, millisecond timescales. So instead of looking at distant optical light coming from stars, they used an existing long-duration observation of Scorpius X-1, which lies close to the ecliptic plane in which Earth's orbit around the Sun lies. Based on the number of rapid dips in brightness on timescales of up to 10 milliseconds, and assuming that the

Box 1 Big ones, small ones, ones of every size

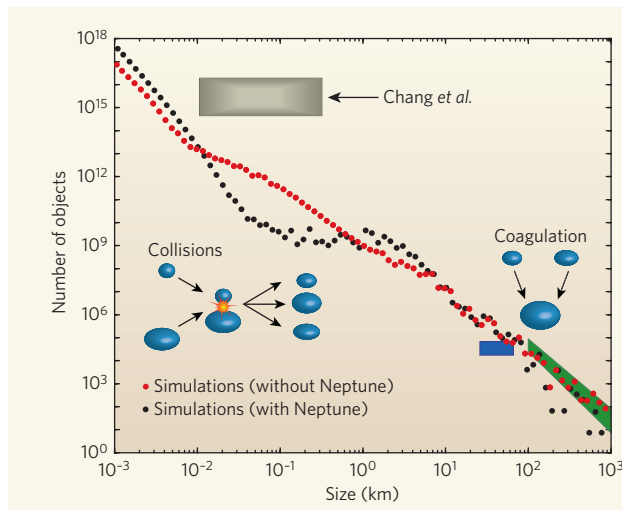
Before Chang and colleagues' results¹, we knew only of the existence of planetoids with sizes greater than a few tens of kilometres in the outer Solar System. These objects were found by the world's largest telescopes looking for faint sources moving slowly relative to the fixed stars. The number of bodies with sizes greater than about 100 kilometres is now estimated to be about 50,000 (regions marked in green and blue).

Chang and colleagues' X-ray occultation technique allows them, for the first time, to see smaller objects. They estimate the total number of small objects of between 10 and 100 metres in size to be around 10^{15} (grey region). That is roughly 1,000 to 100,000 times more than computer models indicate for those sizes

(black and red dotted lines)⁶.

These computer calculations involve two physical processes. At early times, small bodies coagulate to form larger ones; and as the system evolves, fast-moving, smaller bodies collide with slower, larger ones. Neptune, acting as a gravitational stirrer, interacts with these small bodies and was primarily responsible for increasing their speed during its outward migration early in the Solar System's history.

The system of trans-neptunian objects is still evolving today, with collisions slowly grinding small objects to dust. To explain the discrepancy between the computer models and Chang and colleagues' results, models might need to be invoked in which the small bodies move more slowly than the larger



ones to avoid frequent collisions. Another possible solution is to extend the region of small bodies to significantly farther beyond Pluto,

as the observed dips in occultation are not precise enough to establish the distance of each planetesimal independently of its size. **A.C.**

occluding bodies are about as distant as Pluto, the authors estimate that the total number of bodies of between 10 and 100 metres across in the outer Solar System is more than 10^{15} — a quadrillion (Box 1).

That number suggests that there is a 20-metre-sized planetesimal in every sphere of 100,000 kilometres' radius in the outer Solar System. Astronomically speaking, that is extreme overpopulation. The debris field resembles a crowded dance floor where, inevitably, frequent collisions occur. These collisions break up the bodies, grinding them down to smaller objects and eventually to dust particles. Through the interplay of solar radiation pressure and gravitational forces, these particles are slowly removed from the outer Solar System, with some drifting out into space and others spiralling into the Sun. Computer models indicate that there should currently be between a billion (10^9) and a trillion (10^{12}) bodies of 10–100 metres in size⁶, far less than Chang and colleagues' number¹. This discrepancy implies that collisions are less frequent than thought. As faster-moving objects collide more frequently, these small bodies are likely to be moving slowly relative to the larger planetesimals in the disk.

The occultation technique has an inherent problem: a dip does not establish the exact geometry of an event, as a small body nearby would produce a similar dip to that produced by a large object farther away. Fortunately, one can establish the distance to an occulting body independently of its size. This is done by observing the tiny modifications in the shape of a dip as the light beam from the background source diffracts over the surface of the occulting body⁷.

But despite the high X-ray luminosity of the Scorpius X-1 source, it was not bright enough to allow a study of these expected small changes. A dedicated space-based monitoring mission, optimized for precision stellar occultation measurements at millisecond time intervals, could search for diffraction patterns to establish the exact extent of the debris field⁸. Whipple, a mission currently under consideration by NASA, would be able to monitor occultations due to small bodies as far away as the hypothesized Oort cloud at the extreme edge of the Solar System.

The Taiwan–America Occultation Survey⁹ currently operates four robotic optical telescopes in central Taiwan that monitor optical light from stars at 0.2-second intervals. It will soon make a census of the number of kilometre-sized bodies on the outer edges of the Solar System. The cumulative thermal 'black-body' radiation emission from planetesimals could also soon be revealed in exquisite far-infrared images from the European Space Agency's Herschel and Planck missions, which are scheduled for launch in 2008. Meanwhile, theorists must scrutinize their models to explain the dense debris field at the fringes. The reason for its denseness may hide

important clues to the mass and extent of the primordial gas disk from which the Solar System formed.

Asantha Cooray is in the Department of Physics and Astronomy, University of California, Irvine, California 92697, USA.
e-mail: acooray@uci.edu

1. Chang, H.-K. *et al.* *Nature* **442**, 660–663 (2006).

2. Edgeworth, K. E. *J. Br. Astron. Assoc.* **53**, 181–188 (1943).
3. Luu, J. X. & Jewitt, D. C. *Annu. Rev. Astron. Astrophys.* **40**, 63–101 (2002).
4. Elliot, J. L. & Olkin, C. B. *Annu. Rev. Earth Planet. Sci.* **24**, 89–124 (1996).
5. Bailey, M. E. *Nature* **259**, 290–291 (1976).
6. Kenyon, S. J. & Bromley, B. C. *Astron. J.* **128**, 1916–1926 (2004).
7. Roques, F., Moncuquet, M. & Sicardy, B. *Astron. J.* **93**, 1549–1558 (1987).
8. Cooray, A. *Astrophys. J.* **589**, L97–L101 (2003).
9. <http://taos.asiaa.sinica.edu.tw>

NEURODEGENERATIVE DISEASE

Cut to the chase

Lisa M. Ellerby and Harry T. Orr

A family of enzymes called caspases — best known for their involvement in programmed cell death — now seems to be pivotal in the progression of two neurodegenerative diseases.

Caspases are protease enzymes that cut proteins at specific target amino-acid residues. By blocking caspase action on crucial proteins associated with Alzheimer's or Huntington's disease, respectively, Galvan *et al.*¹ and Graham *et al.*² show that the development of symptoms is also blocked in mouse models of the diseases. Although these studies are straightforward in concept, they are each an experimental tour de force, providing compelling *in vivo* evidence that caspases have a prominent role in these devastating conditions.

One of the main characteristics of Alzheimer's disease is the build-up in the brain of 'amyloid plaques' made of clumps of amyloid precursor protein (APP). The amyloid hypothesis posits that a peptide fragment of APP called A β 42 initiates a cascade of events that leads to full-blown Alzheimer's disease. However, APP can be cleaved by caspase at a different site, aspartate amino acid 664 (D664), releasing another toxic peptide, called APP-C31 (refs 3, 4). That suggests that there is more to the story.

Galvan *et al.*¹ used a mouse model of Alzheimer's disease in which the mice express a mutant form of human APP. The authors then mutated the APP further by changing the D664 site to an alanine (D664A). They show convincingly that the mice with the D664A mutation no longer make the APP-C31 peptide. Strikingly, the mice still produce A β 42 and amyloid deposits, but fail to develop the pathological, physiological and behavioural deficits characteristic of Alzheimer's disease.

By contrast, Graham *et al.*² assess the role of caspases in Huntington's disease. This is an inherited disorder associated with a mutated form of huntingtin protein (Htt) that has an expanded stretch of glutamate residues (the polyQ tract). The group had previously genetically engineered a strain of mice expressing mutant Htt with an expanded polyQ tract (called YAC128 mice). These mice develop

many of the features associated with Huntington's disease in humans: cognitive and movement deficits, degeneration of the striatal region of the brain, increased sensitivity to excitotoxicity (the toxic effects on neurons of overactivation) and — significantly — cleavage of Htt with subsequent movement of the protein fragments into the cell nucleus.

Building on evidence that Htt can be digested by caspase-3 and caspase-6, and that these cleavages might be linked to toxicity⁵, Graham *et al.*² generated YAC128 mice that express variations of the polyQ-mutant Htt in which each of the caspase target sites is disrupted by an amino-acid substitution. This yielded mice expressing Htt with the disease-causing polyQ expansion but resistant to cleavage by caspase-6 (which has one possible cleavage site) or caspase-3 (two possible cleavage sites). Mice that entirely lack Htt die before birth, but expression of the caspase-resistant versions of mutant Htt allowed the animals to survive. So, a crucial biological function of Htt, and by inference the overall conformation of Htt, was not altered by the caspase-resistant mutations.

The major finding by Graham *et al.*² was that, in contrast to mice expressing YAC128 Htt or the two forms of caspase-3-resistant Htt, mice expressing caspase-6-resistant Htt failed to develop any Huntington's-like symptoms. In particular, these mice showed no signs of abnormal motor function, neurodegeneration, movement of Htt to the nucleus of striatal cells, or enhanced excitotoxicity.

Galvan *et al.*¹ and Graham *et al.*² demonstrate that, in APP and Htt, specific aspartate residues that are the targets of caspases have central roles in the development of Alzheimer's and Huntington's disease, respectively. Caspases are increasingly being linked to a wide variety of biological processes, but they are best known for being active during programmed cell death. This activation is linked to the cell's response to stress. If the



50 YEARS AGO

"European Brewery Convention"

— The idea of science infiltrating into brewing still frequently produces a reaction from the layman on the grounds that, first, beer has been brewed for more than six thousand years, so there can be little still to learn; second, the less the scientist has to do with it the better because it used to be better than it is (in fact, it is even hinted that nowadays "beer is made from chemicals"); third (triumphantly), brewing is an art and not a science. The first two points are born of ignorance of the real position, and the truth of the third depends on the sense of the word 'art': in so far as it refers to experience and skill in 'know how' it may be correct.

From *Nature* 11 August 1956.

100 YEARS AGO

Poverty and Hereditary Genius; a Criticism of Mr. Francis Galton's Theory of Hereditary Genius —

The criticism which Mr. Constable brings forward in this book is that reputation is not a test of ability, and as Galton's theory of hereditary genius is based on this assumption, it has to be discarded. The statistical evidence given in "Hereditary Genius" has to be explained away, and Mr. Constable attempts to do this by what he calls the "swamping effect of poverty." We quite agree with Mr. Constable that it is harder for a poor man with uninfluential parents to achieve success as a judge than for a rich one with influence, but this does not seem to us to justify Mr. Constable in discarding the conclusions of "Hereditary Genius," for if the social conditions of both parents and offspring are relatively about the same, it seems as if the omission of the ability in poverty-stricken parents and their children is rather like leaving out of account the addition of numbers to both the numerator and denominator of a fraction.

From *Nature* 9 August 1906.

aspartate mutations generated by Galvan *et al.* and Graham *et al.* are indeed having their effect on the neurodegenerative diseases by preventing caspase action, this implies that cleavage of APP and Htt is downstream of some event that activates a cellular response to stress; that is, caspase cleavage and the subsequent pathological events are components of disease progression.

It is notable that the D664A mutation in APP did not affect the generation of A β 42 or amyloid deposition. Perhaps the generation of A β 42 induces a stress signal that results in caspase activation and its pathological consequences. Similarly, caspase-6 cleavage of Htt might also be in response to some stress signal, presumably polyQ-mutant Htt. To understand Huntington's disease fully, this initiating event will need to be identified. Alternatively, caspase-6 might not be the initiating protease for Huntington's disease *in vivo*, in which case identification of this crucial protease would be of paramount concern. It will therefore be essential to examine whether manipulating caspase-6 activity in the brain alters disease progression in YAC128 mice. Interestingly, lowering caspase-6 activity in cell-culture models of Huntington's disease does protect neurons from degeneration⁶.

Caspase-6 has a unique set of substrate specificities that does not overlap with those of other caspases. So selective inhibitors of caspase-6 might block the symptoms of Huntington's disease. Other caspases also interact with APP and Htt, suggesting that blocking such interactions might be beneficial therapeutically^{4,6}. Specific caspase inhibitors are being developed by pharmaceutical companies, but much work needs to be done before we know whether they are suitable for clinical use. The studies of Graham *et al.*¹ and Galvan *et al.*² do, however, validate caspases *in vivo* as potential therapeutic targets for Alzheimer's and Huntington's disease. ■

Lisa M. Ellerby is at the Buck Institute for Age Research, 8001 Redwood Boulevard, Novato, California 94947, USA. Harry T. Orr is at the Institute of Human Genetics, University of Minnesota, 516 Delaware Street SE, Minneapolis, Minnesota 55455-0374, USA.

e-mail: orrrx002@umn.edu

1. Galvan, V. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 7130–7135 (2006).
2. Graham, R. K. *et al.* *Cell* **125**, 1179–1191 (2006).
3. Gervais, F. G. *et al.* *Cell* **97**, 395–406 (2000).
4. Lu, D. C. *et al.* *Nature Med.* **6**, 397–404 (2000).
5. Wellington, C. L. *et al.* *J. Biol. Chem.* **275**, 19831–19838 (2000).
6. Hermel, E. *et al.* *Cell Death Differ.* **11**, 424–438 (2004).

MICROSCOPY

Nanotomography comes of age

David Attwood

The use of X-rays to construct three-dimensional tomographic images is well established in medicine. The same principle is being extended to the nanoscale, bringing us startlingly accurate pictures of tiny objects.

Writing in *Applied Physics Letters*, Yin and colleagues¹ report an X-ray microscopy technique of broad potential for three-dimensional imaging in the physical and life sciences. By tuning high-energy X-rays, the authors manipulate the contributions of specific chemical elements to a series of two-dimensional images. They then use tomographic methods to combine images taken at different incident X-ray angles, allowing internal structures and — given sufficient spectral resolution — chemical bondings to be discerned with a spatial resolution of around 60 nanometres.

In essence, this technique is a nanometre-scale version of medical computed tomographic (CT) imaging of humans. The high spatial resolution of the new system¹ is largely determined by the 50-nm width of the outermost transmitting zone of its zone plate lens. This lens is a circular diffraction grating consisting of alternate transparent and opaque concentric rings^{2,3} and is used to focus the X-ray photons. Higher doses of radiation are required for nanoscale imaging, so radiation-

sensitive samples such as biological tissue can require cryogenic or other 'fixation' techniques that limit structural damage to them.

For their experiments, the authors use photons at wavelengths of between 0.11 and 0.15 nanometres; as a photon's wavelength is inversely proportional to its energy, this is equivalent to photon energies of between 11 and 8 kiloelectronvolts. Such high-energy X-rays are known as hard X-rays. At the 'absorption edges' of a chemical element, photons have enough energy to lift an electron out of a particular atomic orbital, and can therefore be absorbed. Thus, by tuning the X-ray energy to just above or below a prominent absorption edge, the contributions of specific chemical elements to the image can be enhanced or diminished.

Yin *et al.* tested their set-up on a thinned portion of a silicon computer chip containing copper–aluminium wires and an array of tungsten plugs. These plugs are used as electrical interconnects between the typically eight or nine layers of a computer chip. Such plugs

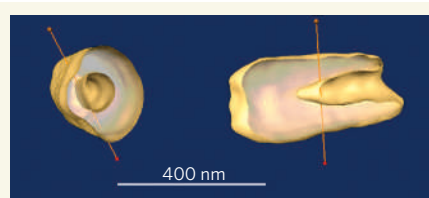


Figure 1 | Plug away. Yin *et al.*¹ use data from 141 two-dimensional X-ray images obtained at 1° angular increments to construct a three-dimensional data set of tungsten plugs used as electrical interconnects between layers of a computer chip. By imaging at an X-ray photon energy of 10.5 keV, just above a tungsten absorption edge, a void of 50–80 nm diameter is clearly seen in the roughly 250 × 500-nm tungsten plugs. (Courtesy of G.-C. Yin.)

are prone to failure through the formation of central voids, known as keyholes, during the deposition process.

The authors tuned their X-rays to a wavelength of 0.118 nm (energy 10.5 keV), just above an absorption edge of tungsten. At this energy, X-rays are hard enough to pass through a silicon layer up to about 50 micrometres thick, ensuring an appropriately contrasted image. Two-dimensional images with incident X-ray angles from –70° to 70° in 1° increments were taken, with an exposure time of about a minute each. Once these data were tomographically reconstructed to a full three-dimensional image, voids in the tungsten plugs could be clearly identified (Fig. 1).

This technique complements existing methods for smaller objects and elements of lower atomic number that use softer, less penetrating X-rays, typically with photon energies below 1 keV and wavelengths of one to several nanometres. The ‘water window’, for example, which is used for probing whole, hydrated biological cells, lies between the absorption edges of carbon (284 eV or 4.4 nm) and oxygen (543 eV or 2.3 nm). In this energy range, water is transparent to about 10 μm thickness, whereas carbon-rich organelles and other subcellular structures are more absorptive, providing a natural contrast mechanism^{2,3}. Recently, detailed three-dimensional maps, also at about 60 nm spatial resolution, of several cryogenically fixed, whole, fully hydrated and unstained biological samples, including individual yeast⁴ and bacterium⁵ cells, have been produced (Fig. 2).

So what’s next in high-resolution X-ray tomography? A new microscope, XM-2, currently being built for the Advanced Light Source at the Lawrence Berkeley National Laboratory in California, will be dedicated entirely to ‘bio-tomography’. Then the three-dimensional distribution of isotopically labelled proteins produced by organisms with normal, mutated or knocked-out gene sets will be able to be studied quantitatively. Fully automated cryogenics and sample rotation will permit statistically significant data sets to be obtained in shorter times. A similar facility in

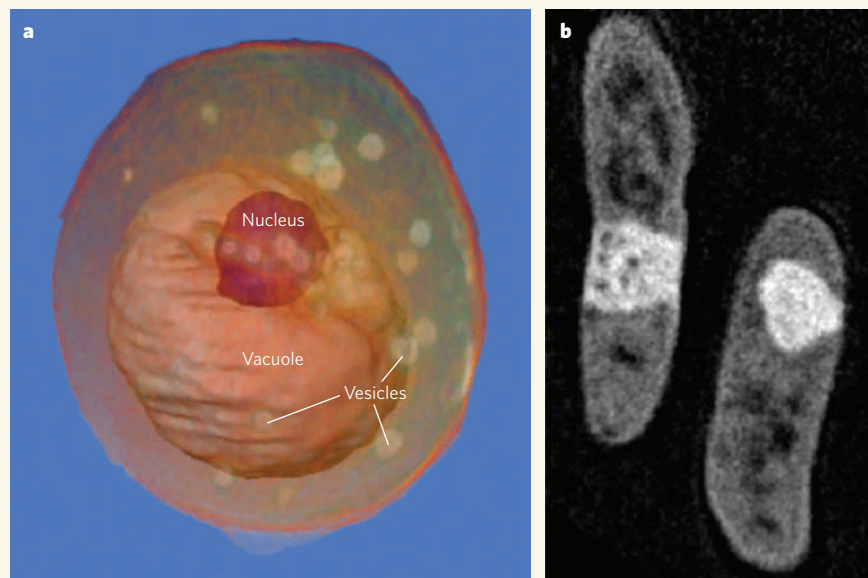


Figure 2 | Nanoscale bio-tomography. **a**, Tomographically reconstructed image of a whole, hydrated yeast cell about 5 μm wide, based on 45 two-dimensional images obtained at 4° angular increments, covering 180°. At the incident photon energy of 517 eV, water is relatively transparent and carbon absorbs, providing a natural contrast mechanism. The cell nucleus, a large vacuole and several lipid-filled vesicles are visible. **b**, Tomographic software can be used to produce images through various planes in an object. Here, a computer-generated section through two *Escherichia coli* bacteria, each about 0.5 μm wide, were obtained from a similar three-dimensional data set from 45 images at 4° intervals and 2.4 nm wavelength. The nature of the large white bands and dark sub-structure is unknown. Typical exposure times are 1–3 seconds per two-dimensional image; it takes about 5 minutes for acquisition of a full data set. (Courtesy of C. Larabell, UCSF and LBNL^{4,5}.)

Berlin⁶ will address cryofixed bio-tomography and a wide range of environmental and nanomagnetic issues. At Taiwan’s National Synchrotron Radiation Research Center, Yin and colleagues are preparing for automated data collection over a full 180° and a broader scientific programme.

The current 60-nm limit on spatial resolution is largely a result of the outer-zone width and thickness of the available zone plates, together with restrictions in the depth of field and the limited angular range in the data sets. All of these should see progress in the near future. Two-dimensional images resolved to 15 nm have been achieved in the soft X-ray region⁷, albeit with a reduced depth of field. For example, the imaging of nanometre-scale magnetic structures formed of elements or compounds containing iron or cobalt, which have absorption edges at 708 eV (1.8 nm) and 778 eV (1.6 nm), respectively, has been demonstrated⁸.

With hard X-rays such as those used by Yin and colleagues¹, 30 nm spatial resolution in two dimensions has been achieved⁹. Higher-energy X-rays generally have the advantage of longer focal lengths, allowing a greater working distance. The shorter wavelengths also allow more effective use of phase-shifting techniques in which refractive effects, rather than absorption, are the key to forming the image. The disadvantage of hard X-rays is the need for thicker zone plates, which are more

difficult to fabricate with the requisite narrow outer zones.

Nonetheless, there has been steady progress on better zone plates for nanoscale imaging throughout the X-ray region, as well as more sophisticated optics and nano-fabrication techniques^{9–12}. The improving infrastructure and increasing availability of well-engineered, automated microscopes for high-throughput studies will ensure that X-ray tomography has a bright future across the materials, environmental and biological sciences. ■

David Attwood is at the University of California, Berkeley, and Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. e-mail: attwood@berkeley.edu

1. Yin, G.-C. *et al.* *Appl. Phys. Lett.* **88**, 241115 (2006).
2. Attwood, D. *Soft X-Rays and Extreme Ultraviolet Radiation: Principles and Applications* (Cambridge Univ. Press, 2000).
3. www.coe.berkeley.edu/AST/sxrev
4. Larabell, C. & Le Gros, M. *Mol. Biol. Cell* **15**, 957–962 (2004).
5. Le Gros, M., McDermott, G. & Larabell, C. *Curr. Opin. Struct. Biol.* **15**, 593–600 (2005).
6. Schneider, G. *et al.* *Surface Rev. Lett.* **9**, 177–184 (2002).
7. Chao, W., Harteneck, B., Liddle, J., Anderson, E. & Attwood, D. *Nature* **435**, 1210–1213 (2005).
8. Fischer, P. *et al.* *Mater. Today* **9**, 26–33 (2006).
9. Tang, M.-T. *et al.* in *Proc. 8th Int. Conf. X-ray Microscopy* (eds Aoki, S., Kagoshima, Y. & Suzuki, Y.) 15 (IPAP, Tokyo, 2006).
10. Fabrizio, E. D. *et al.* *Nature* **401**, 895–898 (1999).
11. Wilhein, T., Kaulich, B. & Susini, J. *Optics Commun.* **193**, 19–26 (2001).
12. Chang, C., Sakdinawat, A., Fischer, P., Anderson, E. & Attwood, D. *Opt. Lett.* **31**, 1564–1566 (2006).

NEUROSCIENCE

Making faces in the brain

James J. DiCarlo

Artificially activating the right neurons at the right time causes visual perception of a face. This new result shows that such neurons directly underlie the recognition of complex objects.

You look up from your desk and instantly recognize the person that has just walked into your office. Unlike the mathematical proof you were busy constructing, recognizing your colleague does not seem to require any mental effort at all. But the apparent ease of this task belies its computational difficulty, and the ability of the primate brain to recognize complex objects, such as faces, in complex environments, such as offices, is a pinnacle of evolutionary achievement. Investigating the brain mechanisms underlying object recognition is also very difficult. However, as they describe on page 692 of this issue, Afraz and colleagues¹ have successfully given us new traction on the problem*.

Although we still have only a rudimentary understanding of how the brain accomplishes visual object recognition, the work of many groups has uncovered a hierarchically ordered series of brain areas likely to support this remarkable feat². One of the most intriguing observations is that neurons at the highest levels of this series respond strongly when their 'preferred' object is in view, often regardless of its exact position, size, contrast, pose and even background clutter. Creating this neuronal response property is the computational crux of object recognition, as even small populations of these neurons can categorize objects with remarkable speed³. We know that this neuronal activity is correlated with object perception by the brain's owner⁴, and even with his or her imagination of objects⁵. Yet there has been no direct evidence that these neurons actually cause subjects to perceive objects. Afraz and colleagues¹ now provide such evidence.

In particular, the authors show that artificial activation of groups of high-level visual neurons that prefer one class of object — faces — causes subjects to tend to perceive faces, even if a face is not present. Working with rhesus monkeys, a well-studied animal model of human vision, the authors focused on the putative highest visual brain area in this species — the inferior temporal cortex (IT) — where neurons respond highly selectively to visually presented faces and other complex objects as outlined above. The researchers used fine microelectrodes to sample the activity of neurons at many locations across IT, and positioned the microelectrode at locations where many neurons responded preferentially to faces — that is, responded with high

activity only when images of faces were shown to the animal. To ask if IT neuronal activity somehow causes perception, the authors injected tiny amounts of electrical current through the microelectrode tip; this method is called microstimulation, and is known to activate neurons within several hundred micrometres⁶.

The prevailing, but previously unsubstantiated view is that IT neurons that respond preferentially to an object are somehow responsible for allowing the subject to perceive and report the presence of that object. So artificial activation of face-preferring neurons by microstimulation might, in theory, lead to face perception.

But how does one ask a monkey what it perceives? To do this, Afraz *et al.*¹ drew on well-established behavioural methods⁷. They first trained each monkey to look left if a face was presented and to look right if a non-face object was presented. Once the animals had mastered this face-detection task, the authors made the task more challenging by degrading the object images with varying amounts of visual noise (see Fig. 1a of their paper¹ on page 692). Averaged over many behavioural trials, this procedure provides a sensitive perceptual assay of the monkey's tendency to report 'seeing' a face rather than a non-face.

Using these methods, the authors showed that, in behavioural trials in which microstimulation was applied to face-preferring neurons, monkeys had a reliably greater tendency to report seeing a face, relative to trials in which no microstimulation was applied. For example, even when the visual stimulus contained no object at all, but only noise, microstimulation increased the chance of the animal 'seeing' a face in that noise. Importantly, these effects were both temporally and spatially specific — they were stronger when microstimulation was applied during a brief time interval in which IT neurons normally respond to visual stimuli, and they did not occur when microstimulation was applied at non-face-preferring IT locations.

Together, these exciting results show that targeted microstimulation of IT induces face perception — probably through the activation of groups of face-selective IT neurons. Because neuronal activity is almost universally believed to underlie conscious perception, it is at one level not surprising that artificial neuronal activation leads to perceptual changes, and this has been shown in other visual areas⁷. However, Afraz and colleagues' study directly

implicates IT neurons as causal in face perception, and also suggests that the millisecond-by-millisecond detail of their activity is not as important as the fact that they are active. Moreover, because the strong microstimulation used activates perhaps thousands of nearby neurons⁶, the study provides firm support for the idea that nearby IT neurons have similar functions⁸. Indeed, it is noteworthy that the study used faces, as they are an ethologically important object class, and face-selective neurons may be clustered together to an unusually strong degree^{9,10}. It remains to be seen if artificially induced perceptual shifts can be obtained for other object classes.

Beyond directly implicating IT in object recognition, such work might eventually allow the creation of visual prosthetics to interact with the brain's high-level object memories. For example, one can imagine even very low-density electrode devices that could activate combinations of IT neuronal clusters to produce awareness of objects (such as a face, car, dog, table, cup, and so on) that are in front of a blind user. However, for deep understanding, the key to progress will be to move from the phenomenology of IT responses to underlying computational mechanisms. In particular, we already know that relatively simple, biologically plausible circuits can 'read out' object identity from IT neurons over behaviourally relevant timescales³. But the central problem is understanding how the brain constructs those IT responses in the first place^{11,12}. That is, how do you recognize that person that just walked into your office? ■

James J. DiCarlo is at the McGovern Institute for Brain Research and the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mail: dicarlo@mit.edu

1. Afraz, S.-R., Kiani, R. & Esteky, H. *Nature* **442**, 692–695 (2006).
2. Rolls, E. T. *Neuron* **27**, 205–218 (2000).
3. Hung, C. P., Kreiman, G., Poggio, T. & DiCarlo, J. J. *Science* **310**, 863–866 (2005).
4. Sheinberg, D. L. & Logothetis, N. K. *Proc. Natl Acad. Sci. USA* **94**, 3408–3413 (1997).
5. Kreiman, G., Koch, C. & Fried, I. *Nature* **408**, 357–361 (2000).
6. Tehovnik, E. J. *J. Neurosci. Meth.* **65**, 1–17 (1996).
7. Salzman, C. D., Murasugi, C. M., Britten, K. H. & Newsome, W. T. *J. Neurosci.* **2**, 2331–2355 (1992).
8. Tanaka, K. *Cereb. Cortex* **13**, 90–99 (2003).
9. Kanwisher, N., McDermott, J. & Chun, M. M. *J. Neurosci.* **17**, 4302–4311 (1997).
10. Tsao, D. Y., Freiwald, W. A., Tootell, R. B. & Livingstone, M. S. *Science* **311**, 670–674 (2006).
11. Kourtzi, Z. & DiCarlo, J. J. *Curr. Opin. Neurobiol.* **16**, 152–158 (2006).
12. Riesenhuber, M. & Poggio, T. *Nature Neurosci.* **2**, 1019–1025 (1999).

*This article and the paper concerned¹ were published online on 26 July 2006.

A second class of chemosensory receptors in the olfactory epithelium

Stephen D. Liberles¹ & Linda B. Buck¹

The mammalian olfactory system detects chemicals sensed as odours as well as social cues that stimulate innate responses. Odorants are detected in the nasal olfactory epithelium by the odorant receptor family, whose ~1,000 members allow the discrimination of a myriad of odorants. Here we report the discovery of a second family of receptors in the mouse olfactory epithelium. Genes encoding these receptors, called 'trace amine-associated receptors' (TAARs), are present in human, mouse and fish. Like odorant receptors, individual mouse TAARs are expressed in unique subsets of neurons dispersed in the epithelium. Notably, at least three mouse TAARs recognize volatile amines found in urine: one detects a compound linked to stress, whereas the other two detect compounds enriched in male versus female urine—one of which is reportedly a pheromone. The evolutionary conservation of the TAAR family suggests a chemosensory function distinct from odorant receptors. Ligands identified for TAARs thus far suggest a function associated with the detection of social cues.

The first step in odour perception is the detection of odorants by G protein-coupled odorant receptors on olfactory sensory neurons (OSNs) in the nasal olfactory epithelium^{1–3}. In response to odorants, OSNs transmit signals to the brain, thereby generating odour perceptions^{2,4}. Each OSN expresses a single functional odorant receptor gene⁵, and OSNs with the same odorant receptor are randomly dispersed within one olfactory epithelial zone^{6,7}. Consistent with their ability to detect and discriminate diverse odorants, mammals have as many as 1,000 different odorant receptors that vary in protein sequence^{8–10} and are used combinatorially to detect different odorants and encode their unique identities⁵. These features of the odorant receptor family would seem to account easily for the odorant recognition abilities of mammals. However, a small percentage of OSNs lack $G_{\alpha_{olf}}$ ¹¹, the G protein through which odorant receptors signal¹², suggesting that they might express another class of chemosensory receptor. In addition, small peptides that bind major histocompatibility complex (MHC) proteins can stimulate some OSNs¹³, suggesting that those OSNs might express a class of receptor that detects peptides rather than small volatile odorants. Finally, although many pheromones are detected in the vomeronasal organ—an olfactory structure with receptors that differ from odorant receptors^{3,14}—responses to some mouse pheromones involve the olfactory epithelium^{15–17}, raising the possibility that the olfactory epithelium also contains a dedicated class of pheromone receptors.

A second family of receptors in the olfactory epithelium

To explore whether there might be other types of chemosensory receptors in the olfactory epithelium, we initiated a search for additional G protein-coupled receptors (GPCRs) expressed by mouse OSNs. In preliminary studies, we found that tissue fixation with glutaraldehyde reveals endogenous β -galactosidase activity present in OSNs, but not in other olfactory epithelial cells (Supplementary Fig. S1). To obtain an enriched population of OSNs, we treated dissociated olfactory epithelial cells with a fluorescent β -galactosidase substrate—fluorescein di-(β -galactopyranoside)¹⁸—and then isolated labelled cells by fluorescence-activated cell sorting. Using RNA from the sorted cells, we prepared complementary DNA,

which we then used in real-time quantitative PCR (qPCR) reactions with primers matching GPCRs not previously implicated in odour, pheromone or taste detection¹⁹. cDNAs encoding individual GPCRs were quantified using standard curves obtained from qPCR reactions with titrations of mouse genomic DNA.

To confirm whether the GPCRs identified by qPCR are actually expressed by OSNs, we used RNA *in situ* hybridization. Initial studies revealed two GPCR genes—*Taar7d* and *Taar9*—that are expressed in small subsets of OSNs. Digoxigenin-labelled antisense RNA probes for each of these genes hybridized to messenger RNA in a small percentage of OSNs that were dispersed in certain olfactory epithelial regions (Fig. 1), an expression pattern similar to that of individual odorant receptor genes^{6,7}. Both of these genes encode members of the trace amine-associated receptor (TAAR) family^{20,21}. On the basis of genome sequence data, this family has 15 members in mouse and 6 in human, and is also found in fish²². TAARs are unrelated to odorant receptors, with their closest relatives being receptors for biogenic amines such as serotonin and dopamine. For example, mouse TAAR1 is 33% identical to the mouse 5-hydroxytryptamine (serotonin) receptor 4, but only 16% identical to the most closely related mouse odorant receptor (OLFR461), and it lacks sequence motifs characteristic of odorant receptors.

To examine whether other TAARs are also expressed in the olfactory epithelium, we used primers specific for each mouse *Taar* gene in qPCR reactions with cDNAs from the olfactory epithelium and other mouse tissues (Fig. 2). These experiments indicated that all mouse *Taar* genes, except *Taar1*, are expressed in the olfactory epithelium. They further showed that the expression levels of individual *Taar* genes in the olfactory epithelium resemble those of odorant receptor genes (Fig. 2).

Although TAARs have been proposed to function as receptors for trace amines (for example, tyramine and octopamine) in the brain, we obtained no evidence for *Taar* gene expression in any tissue—including the brain—apart from the olfactory epithelium, even though we detected high-level expression of genes encoding several biogenic amine receptors in the brain. It could be that TAARs are expressed in small subsets of brain neurons below the detection level

¹Howard Hughes Medical Institute, Basic Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA.

of these assays (~100 copies of mRNA per cell in 60,000 brain cells). The *Taar1* gene is reportedly expressed in mouse and human brain²³ and the *Taar9* gene in human pituitary (and skeletal muscle)²⁴, but qPCR reactions with 50-fold more mouse brain cDNA (detection threshold, ~100 copies of mRNA per cell in 1,200 cells) also failed to reveal the expression of either *Taar1* or *Taar9* in mouse brain (data not shown). Consistent with these results, no expressed sequence tags (ESTs) were listed for any mouse *Taar* gene in the UniGene database of the National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov/>) and, except for two ESTs from stomach (*TAAR1*) and eye (*TAAR2*), all ESTs listed for human *TAAR* genes came from sequence collections derived, in part, from genomic DNA. Together, these data suggest that *TAARs* may be expressed primarily

or exclusively in the olfactory epithelium. However, the expression of some *TAARs* in small subsets of cells in other tissues cannot be excluded.

Zebrafish reportedly has 57 intact *taar* genes²², only one of which—*taar9*—is listed as such at the NCBI. Interestingly, UniGene lists three ESTs for zebrafish *taar9*, all of which are from olfactory rosettes—the location of the fish olfactory epithelium. Using the zebrafish *Taar9* protein to search ESTs translated *in silico* by TBLASTN, we found a large number of zebrafish ESTs encoding related proteins, with 33 of the 37 highest matches being sequences from a zebrafish olfactory epithelium cDNA library (NIH_ZGC_14). The olfactory epithelium ESTs included 24 different *taar* gene sequences, suggesting that numerous different *TAARs* are expressed in the fish olfactory epithelium.

TAAR expression patterns in the olfactory epithelium

We next examined the expression of each mouse *Taar* gene in the olfactory epithelium by RNA *in situ* hybridization²⁵. Similar to results obtained with odorant receptor probes^{6,7}, every *Taar* probe, except a *Taar1* probe, labelled a small subset of OSNs that were scattered in the olfactory epithelium in a seemingly random fashion (Fig. 1). The labelled neurons were confined to certain olfactory epithelial zones, which varied among the *Taar* genes, another feature characteristic of odorant receptor gene expression^{6,7} (Fig. 1). Each *Taar* probe was specific for one *Taar* gene (see below) except the *Taar7* and *Taar8* probes, which are likely to hybridize with all of the highly related members of the *Taar7* (*Taar7a*, *b*, *d*, *e* and *f*) or *Taar8* (*Taar8a*, *b* and *c*) subfamilies, respectively. Expression of each *Taar*, except *Taar1*, was seen in both male and female mice by qPCR as well as RNA *in situ* hybridization. Quantification of OSNs labelled by individual *Taar*

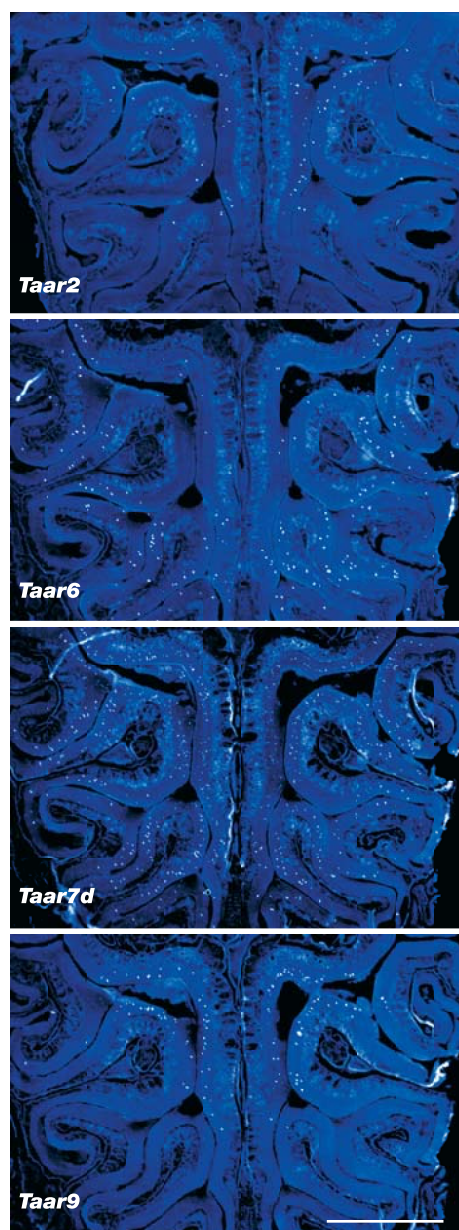


Figure 1 | *Taar* genes are expressed in subsets of olfactory sensory neurons. Digoxigenin-labelled antisense RNA probes for the mouse *Taar* genes indicated were hybridized to coronal sections of mouse olfactory epithelium. Each *Taar* probe hybridized to mRNA in a small percentage of OSNs scattered in selected olfactory epithelial regions. The *Taar7d* probe hybridizes with 5 different *Taar7* subfamily members and labels more OSNs, as well as more olfactory epithelial regions, than the other probes. Scale bar, 1 mm.

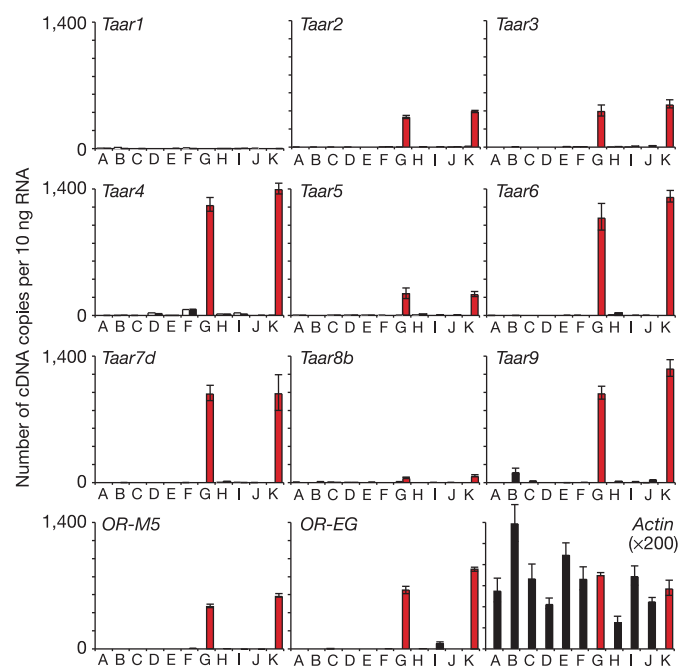


Figure 2 | *Taar* genes are selectively expressed in the olfactory epithelium.

qPCR analysis was performed in triplicate ($\pm$ s.d.) using primers specific for nine mouse *Taar* genes, two mouse odorant receptor genes (*OR-M5* (*Olf139*) and *OR-EG* (*Olf73*)), and the mouse *beta-Actin* gene with cDNAs prepared from different mouse tissues (A, heart; B, spleen; C, intestine; D, liver; E, brain; F, vomeronasal organ; G, olfactory epithelium; H, taste papillae; I, olfactory bulb; J, testis; K, olfactory epithelium (repeat)). cDNAs were prepared with (solid bars) or without (empty bars) reverse transcriptase. All of the *Taar* genes, except *Taar1*, seem to be selectively expressed in the olfactory epithelium (red bars). Note that for actin, the scale of the y-axis is 200 times greater than for the other genes.

probes in every fortieth section along the anterior–posterior axis of the olfactory epithelium (eight sections per probe) gave average counts of 214, 658 and 343 labelled cells for *Taar2*, *Taar6* and *Taar9* probes, respectively. These values were similar to that obtained with an odorant receptor gene probe, *OR-M49* (*Olf11*) (371 labelled cells). With *Taar8* probes, the number of labelled OSNs averaged 212, but the cells were faintly labelled, a result consistent with the lower levels of *Taar8* mRNAs indicated by the qPCR assays (Fig. 2).

To further analyse the expression of TAARs in the olfactory epithelium, we used double-immunofluorescence RNA *in situ* hybridization to compare the expression of pairs of genes²⁵. Double labelling with all possible combinations of *Taar* probes showed that different *Taar* genes are expressed in different OSNs (Fig. 3a, b). Summing all of the pairwise comparisons, 1,097 out of 1,100 OSNs (99.7%) were labelled by one *Taar* probe, but not another.

It is also highly unlikely that TAARs are coexpressed with odorant receptors. In mutant mice that express the odorant receptor *MOR28* (*Olf1507*) in 50–90% of OSNs²⁶, only 0.8% (2 in 250), rather than 50–90%, of neurons that hybridized to *Taar6* or *Taar7f* probes were co-labelled by a *MOR28* probe (Fig. 3d). It is difficult to completely exclude coexpression of odorant receptors and TAARs in non-transgenic olfactory epithelium, because each odorant receptor gene is typically expressed in only ~1 in 1,000 OSNs, or ~1 in 250 in the odorant receptor's expression zone^{6,7}. However, results consistent with the findings in the transgenic mice were obtained when non-transgenic olfactory epithelial sections were double-labelled

with *Taar* probes (*Taar6*, *Taar7e* or a mix of all *Taar* probes) and *OR-M49* or *OR-K20* (*Olf142*) odorant receptor probes, which should hybridize with two and five intact odorant receptor genes (and one odorant receptor pseudogene each), respectively (Fig. 3e). Analyses of *Taar*-positive cells within the same zone as odorant receptor-positive cells showed that 0 out of 174 *Taar*-positive cells were *M49*-positive, and only 6 out of 1,617 *Taar*-positive cells were *K20*-positive. The latter is fivefold fewer than the ~30 out of 1,617 predicted to be positive for odorant receptor expression if each *Taar*-positive neuron randomly selected 1 out of 250 intact odorant receptor genes for expression in the *K20* expression zone. Among odorant receptor-positive neurons, 2,577 out of 2,583 (99.8%) did not hybridize to *Taar* probes.

In contrast, OSNs labelled with probes for *Taar2*, *Taar6*, *Taar7f* or *Taar9* showed the same double-labelling with a $G\alpha_{olf}$ probe (Fig. 3f) as those labelled with an odorant receptor probe. This suggests that, like odorant receptors¹², TAARs may transduce signals by coupling to $G\alpha_{olf}$ and thereby elevate intracellular cAMP levels.

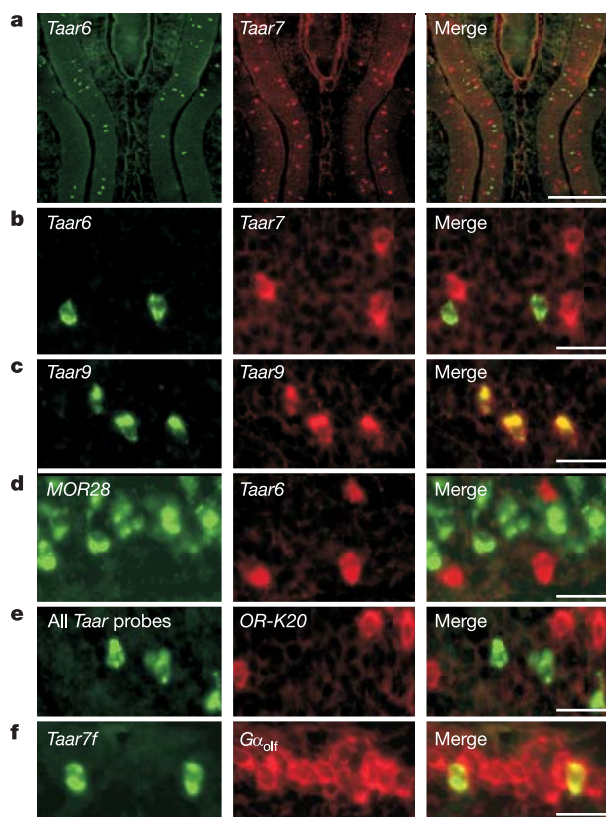


Figure 3 | Each *Taar* gene defines a unique subset of olfactory sensory neurons. **a–f**, The olfactory epithelium expression patterns of different genes were compared using two-colour RNA *in situ* hybridization. Probes for different *Taar* genes labelled different OSNs (**a**, **b**), whereas probes for the same *Taar* labelled the same cells (**c**). *MOR28* and *Taar6* probes labelled different OSNs in *MOR28* transgenic mice (**d**) and, in non-transgenic mice, different OSNs hybridized to an *OR-K20* odorant receptor probe and a mixed *Taar* probe (**e**). OSNs labelled by a *Taar7f* probe were also labelled by a $G\alpha_{olf}$ probe (**f**). Scale bars, 250 μ m (**a**) and 50 μ m (**b–f**).

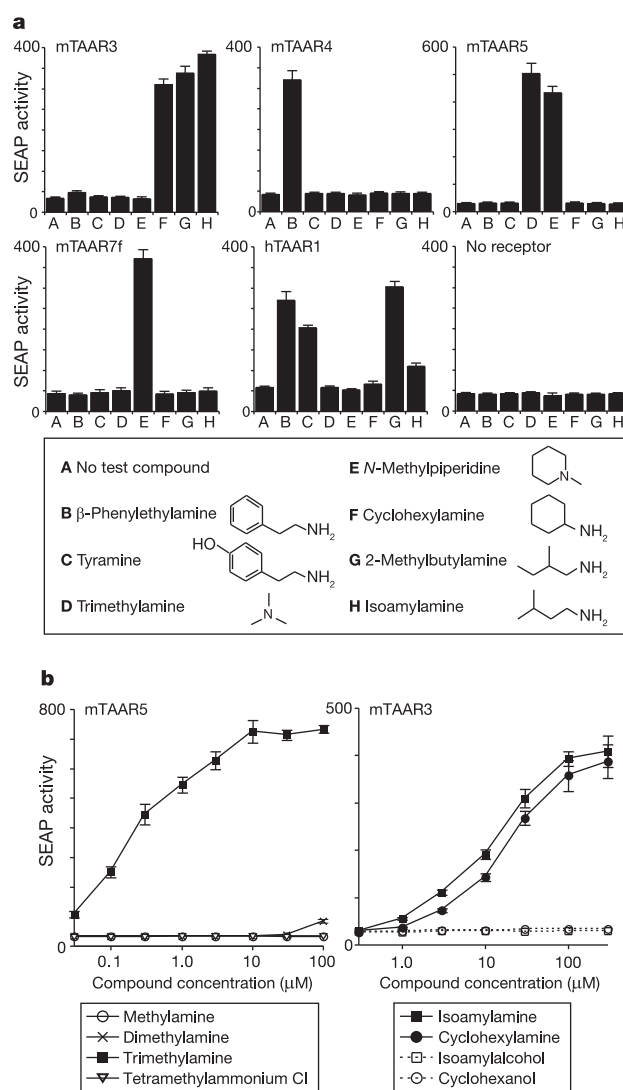


Figure 4 | TAARs recognize volatile amines. **a**, **b**, HEK293 cells were cotransfected with expression vectors encoding TAARs and CRE-SEAP, incubated with different compounds (A–H), and then assayed for SEAP activity (triplicate results $\pm$ s.d.). Different TAARs displayed different ligand specificities (**a**), but all recognized volatile amines. Test compounds were used at 5 μ M (D), 10 μ M (B, C, E) or 20 μ M (F, G, H). Dose–response experiments (**b**) show selectivity of mTAAR5 for a tertiary amine and of mTAAR3 for two primary amines, but not the corresponding alcohols. SEAP activity is expressed in relative fluorescence units.

TAARs recognize volatile amines

The results described above suggested that there are multiple subsets of OSNs that use different TAAR family members rather than odorant receptors for the detection of chemosensory stimuli. We next sought to determine what chemosensory stimuli activate TAARs. To do this, we transfected HEK293 cells with expression vectors encoding individual mouse and human TAARs (referred to as mTAARs and hTAARs, respectively). The cells were cotransfected with the cAMP reporter gene *CRE-SEAP*, which expresses secreted alkaline phosphatase (SEAP) in response to cAMP owing to its cAMP response elements (CRE)²⁷. Following transfection, cells were cultured with potential ligands and then assayed for SEAP activity using the fluorescent SEAP substrate 4-methylumbelliferyl phosphate. Each TAAR was tested with or without an amino-terminal addition of the first 20 amino acids of bovine rhodopsin (a 'rho tag'), a modification that facilitates the cell-surface expression of some odorant receptors in HEK293 cells^{28,29}.

In initial experiments, we tested each TAAR with 30 odorant mixtures containing a total of 210 odorants with diverse structures and perceived odour qualities (each at 2–5 μ M). (See Supplementary Information for a list of all compounds tested.) Consistent with reports that hTAAR1 and rat TAAR4 respond to small organic amines²³, we identified several previously unknown TAAR ligands, all of which were amines. We next tested each TAAR with 81 additional amines, bringing the total number of amines tested to 94. Together, these experiments identified ligands for mTAAR1, mTAAR3, mTAAR4, mTAAR5 and mTAAR7f (Fig. 4a). The mTAARs responded similarly with or without the rho tag, except for mTAAR4, which required the tag for function, and mTAAR3, which only responded without the tag. The identified ligands were

subsequently tested at different concentrations to determine their EC₅₀ values.

In these experiments, mTAAR3 responded to several primary amines, including isoamylamine (EC₅₀ = 10 μ M) (Fig. 4a). In contrast, mTAAR5 and mTAAR7f both responded to tertiary amines: mTAAR5 to trimethylamine (EC₅₀ = 0.3 μ M) and *N*-methylpiperidine, and mTAAR7f to *N*-methylpiperidine (EC₅₀ = 20 μ M) (Fig. 4a). Consistent with previous reports that hTAAR1 and rat TAAR4 both recognize β -phenylethylamine²³, this compound activated hTAAR1, mTAAR1 (EC₅₀ = 0.1 μ M) and mTAAR4 (EC₅₀ = 1 μ M) (Fig. 4a). Slight variations in ligand structure eliminated ligand activity in some cases. For example, mTAAR3 responded to isoamylamine and cyclohexylamine, but not to the corresponding alcohols, isoamylalcohol and cyclohexanol (Fig. 4b). In addition, mTAAR5 was potentially activated by trimethylamine, but not by the related compounds methylamine, dimethylamine and tetramethylammonium chloride, which failed to activate mTAAR5 even at 1,000-fold higher concentrations (Fig. 4b). Although several mTAAR ligands are amino-acid derivatives, mTAARs did not respond to the corresponding amino acids: phenylalanine for β -phenylethylamine (mTAAR1 and mTAAR4); leucine for isoamylamine (mTAAR3); and *N,N*-dimethylglycine for trimethylamine (mTAAR5) (data not shown). No TAAR responses were seen to five mouse pheromones (each at 0.2 μ M) or to an MHC peptide (at 0.5 μ M). However, it cannot be excluded that some TAARs respond to these molecules, but, like many odorant receptors, those TAARs cannot be functionally expressed in HEK293 cells.

Notably, three ligands identified for mTAARs are natural components of mouse urine, a major source of social cues in rodents. mTAAR4 recognizes β -phenylethylamine, a compound whose elevation in urine is correlated with increases in stress and stress responses in both rodents and humans^{30–32}, and both mTAAR3 and mTAAR5 detect compounds (isoamylamine and trimethylamine, respectively) that are enriched in male versus female mouse urine^{33,34}. Furthermore, isoamylamine in male urine is reported to act as a pheromone, accelerating puberty onset in female mice by one criterion³⁴ (although not by another³⁵).

To explore whether other mTAARs might also detect urinary compounds, we treated HEK293 cells expressing each mTAAR with dilutions of mouse urine. Unfortunately, mouse urine generally inhibited responses in the SEAP assay when diluted even 100-fold, making it difficult or impossible to detect TAAR responses to low-abundance urinary compounds. Nonetheless, mTAAR5 responded robustly to highly diluted urine from male mice, but not female mice or humans (Fig. 5a), with the EC₅₀ corresponding to a 30,000-fold dilution of male urine. Higher concentrations of female urine elicited a much weaker response, which was equivalent to that seen with male urine diluted 30-fold more (Fig. 5c). The urine of males did not activate mTAAR5 until the males had reached puberty at about one month of age (Fig. 5b). Urine samples from BALB/c and C57BL/6 male mice were both effective, suggesting that the ligand is not an MHC-linked individuality cue (Fig. 5a). The mTAAR5 activator seemed to be highly volatile as it strongly activated several adjacent wells in a multiwell plate containing mTAAR5-expressing cells. The activator may well be trimethylamine, a highly volatile compound that we identified as an mTAAR5 ligand. If so, on the basis of our data, trimethylamine would be present in male urine at about 9 mM. This is consistent with NMR analysis indicating that trimethylamine is at least as abundant in male mouse urine as creatinine, which is present at 3–5 mM (ref. 33). In addition, NMR analysis indicates that trimethylamine is elevated in male compared with female mouse urine³³, and in mouse urine compared with human urine³⁶. Using mTAAR5, mice could, in principle, determine the gender and sexual status of other mice.

Discussion

In these studies, we identified a second class of chemosensory

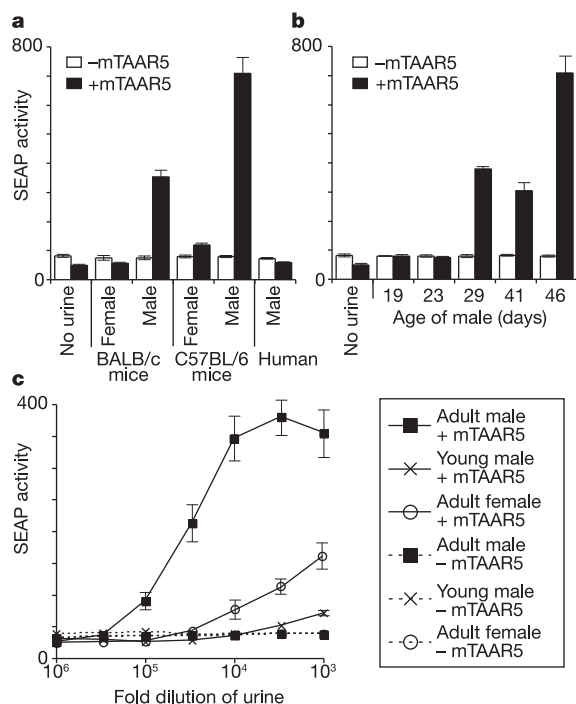


Figure 5 | mTAAR5 is activated by urine from sexually mature male mice. **a–c**, HEK293 cells cotransfected with expression vectors encoding mTAAR5 and CRE-SEAP were incubated with diluted urine from different sources, and then assayed for SEAP activity (triplicate results $\pm$ s.d.). mTAAR5 responded robustly to a 1/30,000 dilution of urine from both BALB/c and C57BL/6 male mice, but not female mice or humans (**a**). The activating substance in male mouse urine was not evident until after puberty (day 29; **b**). Dose–response curves emphasize the differential responsiveness of mTAAR5 to male, female and young (prepubescent) male mouse urine (**c**).

receptors expressed in the nasal olfactory epithelium. These receptors, called TAARs, are expressed in a small subpopulation of neurons that seem to lack odorant receptors, suggesting that these neurons use TAARs rather than odorant receptors to detect chemosensory stimuli. Similar to odorant receptors, different mouse TAARs are expressed in different neurons, and those with the same TAAR are scattered in selected olfactory epithelial regions. OSNs expressing TAARs coexpress $G_{\alpha_{olf}}$, the G protein to which odorant receptors couple, and TAARs can increase cAMP levels when activated by ligands in heterologous cells, suggesting that they activate the same cAMP second-messenger pathway in OSNs as odorant receptors.

These studies show that at least four TAARs expressed in the mouse olfactory epithelium recognize small-molecule amines and, furthermore, that each of these receptors detects a unique set of amine ligands. These findings, together with the relatedness of TAARs to biogenic amine receptors, suggest that TAARs may specifically function as a family of chemosensory receptors for amines. Like biogenic amines such as serotonin, dopamine, adrenaline and histamine, several mTAAR ligands are derivatives of naturally occurring amino acids. β -phenylethylamine (mTAAR4) is decarboxylated phenylalanine and isoamylamine (mTAAR3) is decarboxylated leucine, whereas trimethylamine (mTAAR5) could be derived *in vivo* from *N,N*-dimethylglycine or choline and *N*-methylpiperidine (mTAAR7f) from lysine or homoproline. Future studies should clarify whether all TAAR ligands are biogenic compounds produced in mammals, and, if so, how they are synthesized *in vivo*.

Notably, at least three murine TAARs detect compounds found in mouse urine, an important source of social cues^{14,37}. One detects a chemical (β -phenylethylamine) that is elevated in urine in response to stress, whereas two others detect chemicals (isoamylamine and trimethylamine) that are elevated in male versus female mouse urine. Moreover, one of these compounds (isoamylamine) is reported to act as a male-derived pheromone that accelerates puberty onset in female mice³⁴. In addition, we found that the mouse TAAR that recognizes trimethylamine is activated by extremely dilute mouse urine from sexually mature males, but not females or prepubescent males. Together, these findings suggest that at least some murine TAARs detect social cues that may elicit innate behaviours or physiological responses. This idea is consistent with observations that the olfactory epithelium is involved in some pheromone responses^{14–17}, and that certain pheromone responses may involve dual signals from the olfactory epithelium and vomeronasal organ¹⁶. In this regard, it is interesting that one amine reported to accelerate puberty (isoamylamine) is detected by a TAAR whereas another (isobutylamine) is detected by receptors in the vomeronasal organ³⁸. Future studies should provide information on the ligands of additional TAAR family members and illuminate the functional roles played by individual TAARs.

Families of *Taar* genes are found not only in rodents, but also in humans and fish. Evidence that fish TAARs are also expressed in the olfactory epithelium suggests that TAARs are likely to serve as olfactory receptors in diverse organisms, including humans. The evolutionary conservation of the TAAR family suggests that it may serve a function distinct from the odorant receptor family. The findings presented here suggest that this function may be associated with the detection of social cues such as pheromones.

METHODS

Animals. Adult C57BL/6 mice (Jackson Laboratory) were used except where indicated. *MOR28* transgenic mice²⁶ were generously provided by S. Serizawa and Hitoshi Sakano (Univ. of Tokyo).

OSN isolation. Olfactory epithelial tissue was minced, treated with trypsin/EDTA (Gibco) and 20 units ml⁻¹ DNase (Roche) (10 min, 37 °C), centrifuged (228 g, 5 min, 4 °C), and triturated in phenol-red-free DMEM (Gibco) plus 4% fetal bovine serum (Sigma) (hereafter referred to as 'media'). Dissociated cells were next diluted 1:1 with 2 mM fluorescein di-(β -galactopyranoside), incubated (2 min, 4 °C), diluted tenfold with media containing 1.5 μ M propidium iodide, and then incubated again (2 h at 4 °C). Cells labelled with fluorescein, but

not propidium iodide, were isolated using a Vantage cell sorter (BD Biosciences). **qPCR.** DNase-treated¹ RNA isolated from OSNs or tissues with Trizol (Invitrogen) was used to prepare random-primed cDNA with Superscript III (Invitrogen) (according to the manufacturer's protocols)³⁹. qPCR was conducted using 5- μ l reactions containing cDNA from ~25–50 OSNs or 10 ng tissue RNA with SYBR green indicator and ROX internal reference dye (Invitrogen). Reactions were performed in a Prism 7900HT instrument using SDS 2.2 software (Applied Biosystems). For oligonucleotide primer sequences, see Supplementary Information.

RNA *in situ* hybridization. Digoxigenin- and fluorescein-labelled riboprobes matching coding (or untranslated (*MOR28*)) regions were prepared, hybridized to olfactory epithelial sections, and visualized as described previously²⁵ with minor modifications in blocking and mounting reagents. Hitomi Sakano provided the *MOR28* probe.

TAAR functional assays. *Taar* coding regions were cloned into pcDNA3.1 (Invitrogen) with or without the 5' addition of DNA encoding the N-terminal 20 amino acids of bovine rhodopsin. Functional assays were performed in 96-well plates as described previously⁴⁰ with the following modifications. Each well contained 100,000 HEK293 cells (ATCC) cotransfected (using lipofectamine (Invitrogen)) with 20 ng each of a TAAR plasmid and CRE-SEAP reporter plasmid (BD Biosciences). Cells were incubated for 48 h at 37 °C in serum-free media with or without test compounds, and then for 2 h at 65–70 °C. An aliquot of supernatant from each well was then incubated (2–5 min, ~21 °C) with an equal volume of 1.2 mM 4-methylumbelliferyl phosphate (Sigma) in 2 M diethanolamine bicarbonate, pH 10.0, and fluorescence was measured with a CytoFluor 4000 plate reader (Applied Biosystems). For test compounds, see Supplementary Information.

Received 22 May; accepted 11 July 2006.

Published online 30 July 2006.

- Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
- Kandel, E. R., Schwartz, J. H. & Jessell, T. M. *Principles of Neural Science* (McGraw-Hill, New York, 2000).
- Buck, L. B. The molecular architecture of odor and pheromone sensing in mammals. *Cell* **100**, 611–618 (2000).
- Shepherd, G. M., Chen, W. R. & Greer, C. A. in *The Synaptic Organization of the Brain* (ed. Shepherd, G. M.) 165–216 (Oxford Univ. Press, New York, 2004).
- Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
- Ressler, K. J., Sullivan, S. L. & Buck, L. B. A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell* **73**, 597–609 (1993).
- Vassar, R., Ngai, J. & Axel, R. Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell* **74**, 309–318 (1993).
- Zhang, X. & Firestein, S. The olfactory receptor gene superfamily of the mouse. *Nature Neurosci.* **5**, 124–133 (2002).
- Young, J. M. & Trask, B. J. The sense of smell: genomics of vertebrate odorant receptors. *Hum. Mol. Genet.* **11**, 1153–1160 (2002).
- Godfrey, P. A., Malnic, B. & Buck, L. B. The mouse olfactory receptor gene family. *Proc. Natl Acad. Sci. USA* **101**, 2156–2161 (2004).
- Meyer, M. R., Angele, A., Kremmer, E., Kaupp, U. B. & Muller, F. A cGMP-signaling pathway in a subset of olfactory sensory neurons. *Proc. Natl Acad. Sci. USA* **97**, 10595–10600 (2000).
- Jones, D. T. & Reed, R. R. $G_{\alpha_{olf}}$: an olfactory neuron specific-G protein involved in odorant signal transduction. *Science* **244**, 790–795 (1989).
- Spehr, M. *et al.* Essential role of the main olfactory system in social recognition of major histocompatibility complex peptide ligands. *J. Neurosci.* **26**, 1961–1970 (2006).
- Halpern, M. & Martinez-Marcos, A. Structure and function of the vomeronasal system: an update. *Prog. Neurobiol.* **70**, 245–318 (2003).
- Restrepo, D., Arellano, J., Oliva, A. M., Schaefer, M. L. & Lin, W. Emerging views on the distinct but related roles of the main and accessory olfactory systems in responsiveness to chemosensory signals in mice. *Horm. Behav.* **46**, 247–256 (2004).
- Boehm, U., Zou, Z. & Buck, L. B. Feedback loops link odor and pheromone signaling with reproduction. *Cell* **123**, 683–695 (2005).
- Yoon, H., Enquist, L. W. & Dulac, C. Olfactory inputs to hypothalamic neurons controlling reproduction and fertility. *Cell* **123**, 669–682 (2005).
- Fiering, S. N. *et al.* Improved FACS-Gal: flow cytometric analysis and sorting of viable eukaryotic cells expressing reporter gene constructs. *Cytometry* **12**, 291–301 (1991).
- Vassiliatis, D. K. *et al.* The G protein-coupled receptor repertoires of human and mouse. *Proc. Natl Acad. Sci. USA* **100**, 4903–4908 (2003).
- Lewin, A. H. Receptors of mammalian trace amines. *AAPS J.* **8**, E138–E145 (2006).
- Lindemann, L. & Hoener, M. C. A renaissance in trace amines inspired by a novel GPCR family. *Trends Pharmacol. Sci.* **26**, 274–281 (2005).

22. Gloriam, D. E. *et al.* The repertoire of trace amine G-protein-coupled receptors: large expansion in zebrafish. *Mol. Phylogenet. Evol.* **35**, 470–482 (2005).
23. Borowsky, B. *et al.* Trace amines: identification of a family of mammalian G protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **98**, 8966–8971 (2001).
24. Vanti, W. B. *et al.* Discovery of a null mutation in a human trace amine receptor gene. *Genomics* **82**, 531–536 (2003).
25. Montmayeur, J. P., Liberles, S. D., Matsunami, H. & Buck, L. B. A candidate taste receptor gene near a sweet taste locus. *Nature Neurosci.* **4**, 492–498 (2001).
26. Serizawa, S. *et al.* Negative feedback regulation ensures the one receptor–one olfactory neuron rule in mouse. *Science* **302**, 2088–2094 (2003).
27. Durocher, Y. *et al.* A reporter gene assay for high-throughput screening of G-protein-coupled receptors stably or transiently expressed in HEK293 EBNA cells grown in suspension culture. *Anal. Biochem.* **284**, 316–326 (2000).
28. Krautwurst, D., Yau, K. W. & Reed, R. R. Identification of ligands for olfactory receptors by functional expression of a receptor library. *Cell* **95**, 917–926 (1998).
29. Kajiya, K. *et al.* Molecular bases of odor discrimination: reconstitution of olfactory receptors that recognize overlapping sets of odorants. *J. Neurosci.* **21**, 6018–6025 (2001).
30. Paulos, M. A. & Tessel, R. E. Excretion of β -phenethylamine is elevated in humans after profound stress. *Science* **215**, 1127–1129 (1982).
31. Snoddy, A. M., Heckathorn, D. & Tessel, R. E. Cold-restraint stress and urinary endogenous β -phenylethylamine excretion in rats. *Pharmacol. Biochem. Behav.* **22**, 497–500 (1985).
32. Grimsby, J. *et al.* Increased stress response and β -phenylethylamine in MAOB-deficient mice. *Nature Genet.* **17**, 206–210 (1997).
33. Gavanagh McKee, C. L., Wilson, I. D. & Nicholson, J. K. Metabolic phenotyping of nude and normal (Alpk:ApfCD, C57BL10J) mice. *J. Proteome Res.* **5**, 378–384 (2006).
34. Nishimura, K., Utsumi, K., Yuhara, M., Fujitani, Y. & Iritani, A. Identification of puberty-accelerating pheromones in male mouse urine. *J. Exp. Zool.* **251**, 300–305 (1989).
35. Price, M. A. & Vandenbergh, J. G. Analysis of puberty-accelerating pheromones. *J. Exp. Zool.* **264**, 42–45 (1992).
36. Lee, M. B., Storer, M. K., Blunt, J. W. & Lever, M. Validation of ^1H NMR spectroscopy as an analytical tool for methylamine metabolites in urine. *Clin. Chim. Acta* **365**, 264–269 (2006).
37. Novotny, M. V. Pheromones, binding proteins and receptor responses in rodents. *Biochem. Soc. Trans.* **31**, 117–122 (2003).
38. Del Punta, K. *et al.* Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes. *Nature* **419**, 70–74 (2002).
39. Wittwer, C. T., Herrmann, M. G., Moss, A. A. & Rasmussen, R. P. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques* **22**, 130–139 (1997).
40. Clipstone, N. A. & Crabtree, G. R. Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. *Nature* **357**, 695–697 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Serizawa and Hitoshi Sakano for generously providing MOR28 transgenic mice. We also thank K. Wilson and R. Childs for technical assistance, and members of the Buck laboratory for helpful comments. This project was supported by the Howard Hughes Medical Institute and by grants from the National Institutes of Health (NIDCD).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.B. (lbuck@fhcrc.org).

Card9 controls a non-TLR signalling pathway for innate anti-fungal immunity

Olaf Gross¹, Andreas Gewies¹, Katrin Finger¹, Martin Schäfer², Tim Sparwasser², Christian Peschel¹, Irmgard Förster^{2†} & Jürgen Ruland¹

Fungal infections are increasing worldwide due to the marked rise in immunodeficiencies including AIDS; however, immune responses to fungi are poorly understood. Dectin-1 is the major mammalian pattern recognition receptor for the fungal component zymosan. Dectin-1 represents the prototype of innate non-Toll-like receptors (TLRs) containing immunoreceptor tyrosine-based activation motifs (ITAMs) related to those of adaptive antigen receptors. Here we identify Card9 as a key transducer of Dectin-1 signalling. Although being dispensable for TLR/MyD88-induced responses, Card9 controls Dectin-1-mediated myeloid cell activation, cytokine production and innate anti-fungal immunity. Card9 couples to Bcl10 and regulates Bcl10–Malt1-mediated NF- κ B activation induced by zymosan. Yet, Card9 is dispensable for antigen receptor signalling that uses Carma1 as a link to Bcl10–Malt1. Thus, our results define a novel innate immune pathway and indicate that evolutionarily distinct ITAM receptors in innate and adaptive immune cells use diverse adaptor proteins to engage selectively the conserved Bcl10–Malt1 module.

Pathogen recognition by immune cell receptors is essential for host defence against microorganisms. Whereas the highly diverse T- and B-cell receptors (TCRs and BCRs) for adaptive immunity are generated through somatic rearrangement of antigen receptor genes, the evolutionarily ancient pattern recognition receptors (PRRs) of the innate immune system are germline encoded and control the first line of defence. These PRRs include the archetypical TLRs as well as non-TLRs like intracellular NOD proteins, lectin receptors and others^{1–3}. Upon ligand binding, both innate and adaptive immunoreceptors engage intracellular signalling pathways that converge on conserved core transcription factors for cell activation. To mediate specificity, each receptor system uses distinct sets of signalling molecules that relay proximal events to downstream effectors. Although substantial progress has been made in deciphering TCR, BCR and TLR signal transduction^{1,4}, the inflammatory pathways engaged by non-TLR PRRs remain poorly defined.

One key transcription factor for both innate and adaptive immunity is NF- κ B⁵. Almost all TLRs signal via MyD88 for NF- κ B activation and subsequent inflammatory cytokine production¹. In contrast, the T- and B-cell antigen receptors activate proximal Src and Syk tyrosine kinases and then engage the caspase recruitment domain (CARD) proteins Carma1 and Bcl10 for downstream signalling. Together with Malt1, the Carma1–Bcl10 complex mediates protein kinase C (PKC)-controlled NF- κ B induction for lymphocyte activation and differentiation⁶. The poorly characterized Card9 protein is structurally related to Carma1 and is expressed in a variety of tissues including peripheral blood lymphocytes and the spleen⁷. Similar to Carma1 and the Carma family members 2 and 3, Card9 contains a CARD that can bind to the Bcl10 CARD as well as a coiled-coil region. However, Card9 lacks the carboxy-terminal MAGUK domain that characterizes Carma proteins and the Carma linker region that regulates PKC-dependent cell activation^{8,9}. As such, Card9 has been suggested to have a modulatory role in the

Carma1/Bcl10-dependent pathway of lymphocyte differentiation and activation, possibly by interfering with the Carma1–Bcl10 interaction¹⁰. Here we report that Card9 is surprisingly not involved in TCR or BCR signalling. Instead, Card9 operates as a key adaptor for non-TLR PRR signal transduction, which is required to link Dectin-1/Syk activation to Bcl10–Malt1-dependent NF- κ B activation for innate anti-fungal immunity.

Normal adaptive immunity without Card9

To investigate the functions of Card9 we generated Card9-deficient (*Card9*^{−/−}) mice (Supplementary Fig. 1) and first analysed the requirement of Card9 for adaptive immunity. Flow cytometric analysis revealed normal development of all tested *Card9*^{−/−} T- and B-cell subsets including regulatory T cells, recirculating B cells, mature follicular and marginal zone B cells as well as B1 B cells (Supplementary Fig. 2 and data not shown). Furthermore, no significant differences in basal immunoglobulin levels were detected between wild type and *Card9*^{−/−} mice (Fig. 1a). Upon immunization with either the T-cell-independent antigen 2,4,6-trinitrophenyl (TNP) conjugated to Ficoll or the T-cell-dependent antigen 4-hydroxy-3-nitrophenylacetyl (NP) conjugated to ovalbumin we detected normal IgM and IgG responses in *Card9*^{−/−} animals (Fig. 1b). To study directly antigen-receptor-mediated lymphocyte activation, we stimulated purified T and B cells through their respective antigen receptors either alone or together with co-receptor ligation or with phorbol ester and calcium ionophore (Iono, Fig. 1c). Normal viability (data not shown) and normal signal-induced proliferation was detected in *Card9*^{−/−} lymphocytes. We also found normal TCR- or BCR-induced NF- κ B activation in *Card9*^{−/−} cells (data not shown). Thus, we conclude that Card9 is unlikely to be critically involved in the Carma1–Bcl10 signalling pathway that mediates T- and B-cell differentiation and activation for adaptive immunity.

¹III. Medizinische Klinik, Klinikum rechts der Isar, Technische Universität München, Ismaninger Str. 22, 81675 Munich, Germany. ²Institut für Medizinische Mikrobiologie, Immunologie und Hygiene, Technische Universität München, Trogerstr. 30, 81675 Munich, Germany. [†]Present address: Institut für Umweltmedizinische Forschung, University of Düsseldorf, Auf'm Hennekamp 50, 40225 Düsseldorf, Germany.

Card9 controls zymosan-induced cell activation

Because Card9 is also expressed in myeloid cells⁷ (Fig. 2a) we reasoned that Card9 might have a role in innate immunity. Therefore we generated bone-marrow-derived dendritic cells (BMDCs) from *Card9*^{-/-} mice and stimulated these with various TLR ligands as well as ligands for non-TLR PRRs (Fig. 2b) and measured ligand-induced production of tumour necrosis factor- α (TNF- α). No reproducible differences were detected between wild type and *Card9*^{-/-} BMDCs in response to the TLR ligands Pam₃CSK₄ (TLR2), lipopolysaccharide (LPS; TLR4), R848 (TLR7), CpG DNA (TLR9) or upon stimulation with the Nod2 agonist muramyl dipeptide (MDP) alone or in

combination with LPS. In contrast, TNF- α production induced by zymosan is severely defective in the absence of Card9.

Zymosan is a major yeast cell wall component that is principally composed of β -glucans and additionally contains α -mannan and mannoproteins¹¹. The main mammalian PRR for zymosan is the β -glucan receptor Dectin-1 (ref. 12), yet zymosan also co-stimulates TLR2 (ref. 13). To characterize further the requirement of Card9 in zymosan-induced dendritic cell activation, we obtained dose-response curves for TNF- α , interleukin (IL)-6 and IL-2 production. A strong Card9 gene dosage dependency was observed for the production of all three cytokines (Fig. 2c). Comparable results were obtained with zymosan preparations from two independent sources (data not shown). In contrast, LPS stimulation induced regular dose-dependent TNF- α and IL-6 synthesis in *Card9*^{-/-} BMDCs (Fig. 2d), whereas LPS failed to induce IL-2 production in both *Card9*^{-/-} and control cells (data not shown).

It is believed that the recognition of zymosan or β -glucans triggers immune responses that are primarily designed for the control of fungal pathogens¹⁴. Therefore, we next studied the role of Card9 in immune responses to whole fungal cells using *Candida albicans* as a model¹⁵. Consistent with a role of Card9 in zymosan recognition signalling, *Card9*^{-/-} BMDCs exhibit severe defects in *C. albicans*-induced TNF- α , IL-6 and IL-2 production (Fig. 3a). To determine the relevance of these findings *in vivo* we injected live *C. albicans* into *Card9*^{-/-} mice and heterozygous littermates and compared their

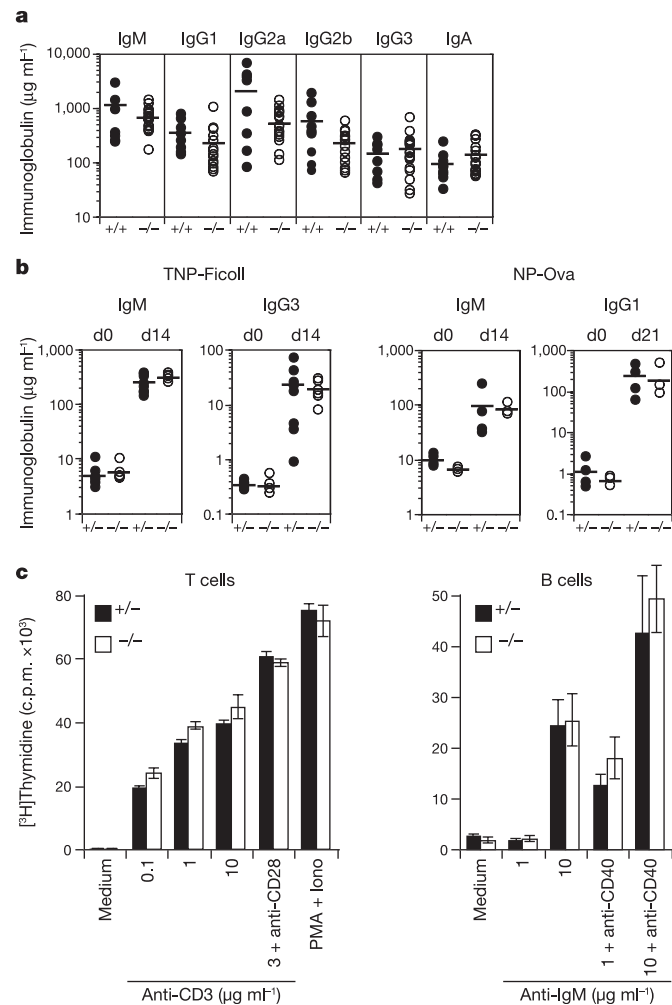


Figure 1 | Regular adaptive immunity in *Card9*-deficient mice. **a**, Basal immunoglobulin concentrations. Immunoglobulin isotypes in sera of 8–12-week-old wild type (+/+, $n = 9$) and *Card9*^{-/-} ($n = 18$) mice are shown. **b**, Regular antibody responses. Left: control (+/+, filled symbols, $n = 8$) and *Card9*^{-/-} mice (open symbols, $n = 6$) were immunized with TNP-Ficoll for T-cell-independent immune responses. Concentrations of TNP-specific IgM and IgG3 were determined before and after immunization. Right: for T-cell-dependent antibody responses, control (+/+, filled symbols, $n = 4$) and *Card9*^{-/-} mice (open symbols, $n = 3$) were injected with NP-Ova absorbed on alum. NP-specific IgM and IgG3 concentrations were measured before and after immunization. Results are representative for two independent experiments. **c**, Signal-induced lymphocyte proliferation. CD4⁺ T cells and B220⁺ B cells were stimulated for 48 h with agonistic antibodies (anti-IgM or anti-CD3, as indicated; anti-CD28, 2 $\mu\text{g ml}^{-1}$; anti-CD40, 5 $\mu\text{g ml}^{-1}$) or PMA plus calcium ionophore (Iono) (10 ng ml^{-1} each) and assessed for proliferation by measuring [3H]thymidine incorporation. Data are representative of four independent experiments with a total of $n = 18$ *Card9*^{-/-} and $n = 18$ control mice. Indicated values are means $\pm$ s.d. of triplicates.

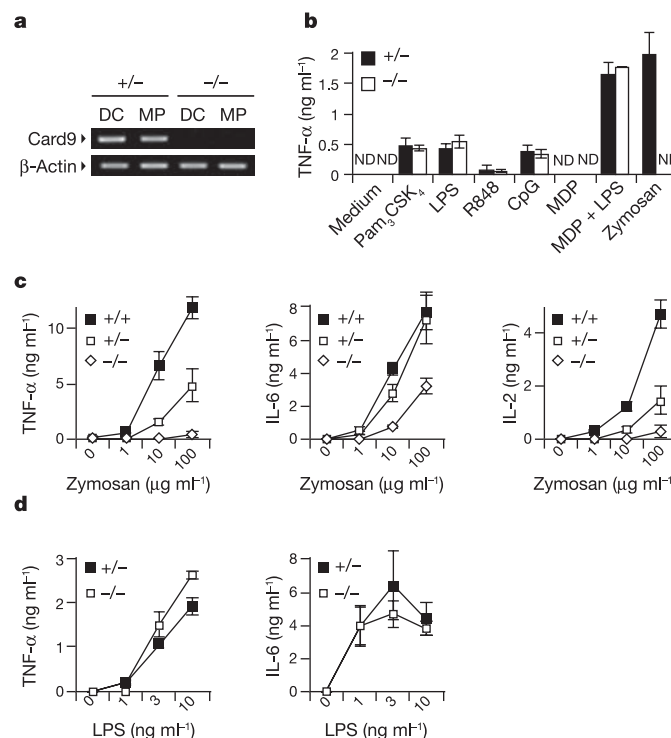


Figure 2 | Impaired zymosan-induced cytokine production in *Card9*^{-/-} BMDCs. **a**, RT-PCR analysis of *Card9* expression in BMDCs (DC) and BMDMs (MP) from *Card9*^{-/-} mice. **b**, TNF- α production in BMDCs from *Card9*^{+/+} or *Card9*^{-/-} mice that were left unstimulated (Medium) or stimulated with TLR ligands Pam₃CSK₄ (200 ng ml^{-1}), LPS (100 ng ml^{-1}), R848 (200 ng ml^{-1}), CpG (1 μM), MDP (10 $\mu\text{g ml}^{-1}$) or zymosan (3 $\mu\text{g ml}^{-1}$) for 24 h. ND, not detected. Data are means $\pm$ s.d. of triplicates and representative of three independent experiments. **c**, Dose-dependent TNF- α , IL-6 or IL-2 production from *Card9*^{+/+}, *Card9*^{+/-} or *Card9*^{-/-} BMDCs after 24 h of zymosan stimulation. Data are means $\pm$ s.d. of triplicate samples and representative of three independent experiments. **d**, Dose-dependent TNF- α or IL-6 production from *Card9*^{+/+} or *Card9*^{-/-} BMDCs after 24 h of LPS stimulation. Data are means $\pm$ s.d. of triplicate samples and representative of three independent experiments.

susceptibility to infection (Fig. 3b). All *Card9*-deficient animals died within 5 days after infection whereas more than 50% of the control mice survived for more than 12 days. In another set of experiments we killed the animals 4 days after injection and assessed intravital *C. albicans* growth. It was macroscopically apparent that the kidneys of all *Card9*^{-/-} animals were massively infiltrated by the fungus (Fig. 3c). Compared with control mice, we detected a 100-fold higher load of *C. albicans* in the kidneys and livers of *Card9*^{-/-} mice and additionally observed high *C. albicans* titres in their lungs (Fig. 3d). In contrast, *Card9*^{-/-} mice regularly cleared the bacterial pathogen *Staphylococcus aureus* (Fig. 3e). Thus, *Card9* is critically required for zymosan-induced dendritic cell activation and innate anti-fungal immunity.

Dectin-1/Syk signalling depends on *Card9*

Zymosan stimulates Dectin-1 and TLR2, which cooperate for full cell activation^{12,13,16–18}. Dectin-1 represents the prototype of non-TLR PRRs that contain ITAMs related to those of antigen receptors or Fc receptors¹⁹. Zymosan-induced Dectin-1 signalling depends on an intact ITAM and downstream activation of Syk, whereas zymosan-induced TLR signalling is transduced via MyD88 (ref. 17). To investigate selectively the involvement of *Card9* in the Dectin-1-independent TLR2 pathway, we used two independent pure TLR2 agonists, Pam₃CSK₄ and peptidoglycan (PGN), for BMDC stimulation^{20,21}. Consistent with the normal TLR2-dependent clearance of *S. aureus*²² in *Card9*^{-/-} mice, both ligands induced normal TNF- α

and IL-6 production (Fig. 4a and data not shown) in the absence of *Card9*.

Because there is no single agonistic Dectin-1 ligand available, it is more intricate to study Dectin-1/Syk signalling without considering TLR2 co-stimulation. However, recent work has shown that the Dectin-1/Syk axis selectively controls zymosan-induced IL-10 production entirely independently of TLR/MyD88 signalling¹⁷. We thus used zymosan-induced IL-10 production as a specific readout for the Dectin-1/Syk pathway. In addition to *Card9*^{-/-} and wild-type BMDCs we included BMDCs from *Myd88*^{-/-} mice²³ as well as wild-type BMDCs that were pre-incubated with the small molecule Syk inhibitor piceatannol as selective controls for TLR/MyD88 and Dectin-1/Syk signalling (Fig. 4b). Consistent with published data¹⁷ zymosan induces strong IL-10 production in wild-type BMDCs that is MyD88-independent but robustly blocked by inhibiting Syk. Deletion of *Card9* also completely abolishes zymosan-induced IL-10 synthesis, indicating a function for *Card9* in Dectin-1/Syk-dependent, TLR/MyD88-independent events. Compared to zymosan, the TLR ligands LPS and CpG induce lower, but regular amounts of IL-10 in *Card9*^{-/-} BMDCs (Fig. 4c). We next studied zymosan-induced production of TNF- α , IL-6, IL-2 and IL-12 in wild-type BMDCs with or without Syk inhibitor and in *Card9*^{-/-} BMDCs (Fig. 4d). Zymosan-induced production of TNF- α and IL-2 is highly dependent on Syk and *Card9*. However, consistent with previous data the production of IL-12 (ref. 17) and IL-6 is at least in part independent of Syk and also independent of *Card9* function. Thus, the effects of Syk inhibition or *Card9* deletion on zymosan-induced cytokine production are in each case comparable, further suggesting an essential function of *Card9* in Dectin-1/Syk signalling.

Card9 relays Dectin-1/Syk signals to NF- κ B

To identify the molecular role of *Card9* in Dectin-1/Syk-mediated cell activation we determined Dectin-1 expression and Dectin-1-triggered Syk phosphorylation in *Card9*^{-/-} BMDCs. The receptor is regularly expressed (data not shown) and normally activates Syk (Fig. 4e). Syk activation does not only control cytokine production but also regulates zymosan phagocytosis¹⁷. To quantify zymosan internalization we incubated BMDCs with or without Syk inhibitor together with fluorescently labelled zymosan (Fig. 4f). Using fluorescence-activated cell sorting (FACS), we observed normal Syk-dependent phagocytosis in *Card9*^{-/-} BMDCs, demonstrating that certain pathways controlled by Dectin-1/Syk do not depend on *Card9*.

Because overexpression studies have indicated that *Card9* can induce NF- κ B activation⁷, we probed NF- κ B activation by monitoring RelA nuclear translocation in zymosan- or LPS-stimulated BMDCs (Fig. 4g). TLR4 stimulation induced normal RelA nuclear translocation in *Card9*^{-/-} BMDCs. Microscopy also demonstrated that *Card9*^{-/-} BMDCs internalized zymosan normally. However, these cells exhibit a substantial impairment in zymosan-induced NF- κ B activation, although residual NF- κ B induction presumably via TLR2 could be observed. In addition, we performed electromobility shift assays using nuclear extracts from bone-marrow-derived macrophages (BMDMs) that were stimulated with zymosan or PGN, LPS or R848 (Fig. 4h). Whereas TLR ligation induced normal NF- κ B DNA binding, zymosan-induced NF- κ B activation was diminished in the absence of *Card9*. Thus, *Card9* is specifically required to transduce zymosan-activated Dectin-1 signals to NF- κ B.

Card9 signals via Bcl10 and Malt1

As *Card9* can interact with Bcl10 (ref. 7), we studied whether the two proteins functionally cooperate (Fig. 5). Consistent with previous results⁷ single overexpression of *Card9* or Bcl10 weakly activates NF- κ B. However, co-overexpression of both proteins induced more than additive NF- κ B induction, indicating synergism (Fig. 5a). To establish a hierarchy between *Card9* and Bcl10 we introduced *Card9* into wild type or Bcl10-deficient murine embryonic fibroblasts²⁴ (Fig. 5b). In parallel we overexpressed the Bcl10 upstream activator

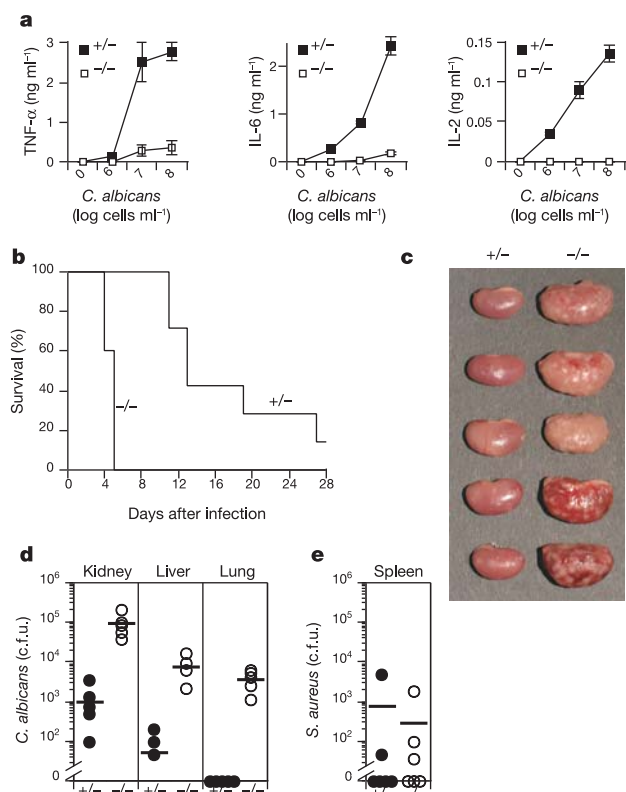


Figure 3 | Impaired immune responses to *Candida albicans*. **a**, *C. albicans*-induced TNF- α , IL-6 or IL-2 production in *Card9*^{-/-} BMDCs. Cells were stimulated for 24 h. Results are means $\pm$ s.d. of triplicates. **b**, *Card9*^{+/-} and *Card9*^{-/-} mice were infected intravenously with 1.5×10^5 colony-forming units (c.f.u.) *C. albicans* and monitored daily for survival. One out of three independent experiments is shown. **c**, **d**, *Card9*^{+/-} and *Card9*^{-/-} mice were infected intravenously with 3×10^4 c.f.u. *C. albicans*. Four days later the left kidneys were inspected (**c**) and *C. albicans* titres were determined in one kidney, the liver and the lung (**d**). **e**, *Card9*^{+/-} and *Card9*^{-/-} mice were infected intravenously with 5×10^5 c.f.u. *S. aureus*. Five days later *S. aureus* titres were determined in the spleen.

Carma3 (ref. 25) and the Bcl10-independent NF- κ B inducer Nod1. Nod1 induced equal NF- κ B activation in wild type and *Bcl10*^{-/-} cells, whereas Carma3 activated NF- κ B only in *Bcl10*^{+/-} cells. Similarly, Card9 induced NF- κ B activation only in wild-type cells, indicating a requirement for Bcl10 in Card9-mediated NF- κ B signalling. To investigate directly the role of a Card9-Bcl10 complex in Dectin-1 signalling, we introduced Dectin-1 and a combination of Card9 and Bcl10 either alone or together into 293 cells (which express low levels of endogenous Syk²⁶) and stimulated these with zymosan (Fig. 5c). Overexpression of Dectin-1 with or without zymosan treatment did not induce NF- κ B activity. In addition, zymosan treatment did not increase the Card9-Bcl10-induced NF- κ B luciferase signal without Dectin-1 expression. However, co-expression of Dectin-1 and Card9-Bcl10 conferred 293 cells with the ability to respond to zymosan, resulting in a substantial additional increase in NF- κ B reporter activity upon zymosan treatment.

To analyse the function of Bcl10 in the Card9-dependent Dectin-1/Syk pathway, we stimulated BMDCs from *Bcl10*^{-/-} mice²⁴ with zymosan or TLR ligands. Because signals downstream of Bcl10 can involve Malt1 we also examined BMDCs from *Malt1*^{-/-} animals²⁷. Bcl10- and Malt1-deficient BMDCs produced normal amounts of TNF- α in response to PGN, LPS and CpG (Fig. 5d). In addition, zymosan phagocytosis was not affected by Bcl10 or Malt1 deficiency (data not shown). However, zymosan-induced cytokine production was severely defective in both *Bcl10*^{-/-} and *Malt1*^{-/-} BMDCs (Fig. 5e). Next, we stimulated *Bcl10*^{-/-} and *Malt1*^{-/-}

BMDCs with whole fungal cells of *C. albicans* (Supplementary Fig. 3a). Severe defects in TNF- α , IL-2 and IL-10 production were detected. Finally, we investigated the roles of Bcl10 and Malt1 in zymosan-induced NF- κ B signalling (Supplementary Fig. 3b). As before, zymosan induced robust nuclear translocation of RelA in wild-type cells. However, similar to *Card9*^{-/-} BMDCs, both Bcl10- and Malt1-deficient BMDCs showed an impairment in zymosan-induced NF- κ B activation, whereas they exhibited normal nuclear translocation of RelA upon LPS stimulation.

Bcl10-Malt1 signalling in innate and adaptive immunity

Collectively, our results identify a novel signalling pathway for innate immunity. Whereas Card9 is dispensable for adaptive antigen receptor signalling as well as for Bcl10-dependent Fc ϵ RI signalling²⁸ (not shown), our data support the hypothesis that Card9 operates upstream of Bcl10 to transduce, together with Bcl10 and Malt1, non-TLR signals for NF- κ B activation and cytokine production. For a model see Fig. 5f. In contrast to immunity directed against viruses and bacteria, far less is known about innate responses to fungal infections. Yet, the incidence of these infections is rapidly growing with the worldwide increase in the number of immunocompromised patients suffering from AIDS or undergoing immunosuppressive therapies. To potentially manipulate anti-fungal immunity it is necessary to understand the pathways controlling these responses. Dectin-1 was recently discovered as the key non-TLR PRR for fungal β -glucan detection¹⁹. This receptor uses an intracellular YxxL ITAM

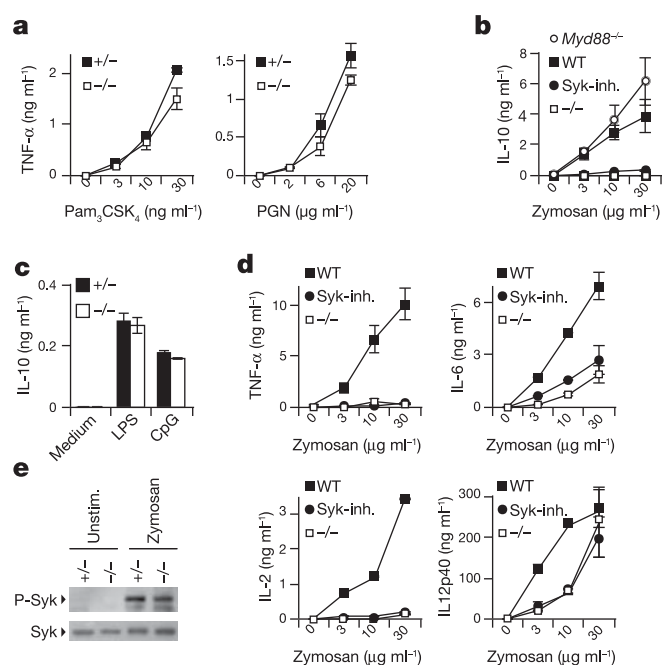
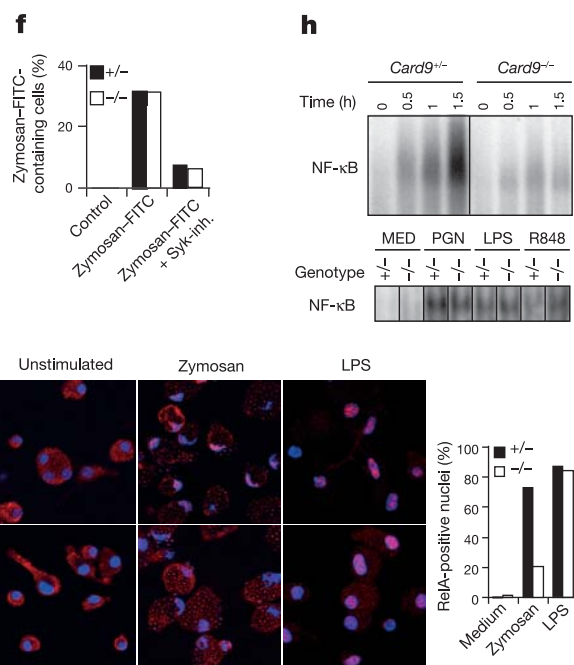


Figure 4 | Impaired Dectin-1/Syk signalling in *Card9*^{-/-} cells. **a**, Normal TLR2-induced TNF- α production. BMDCs from *Card9*^{+/-} or *Card9*^{-/-} mice were stimulated for 24 h with the selective TLR2 ligands Pam₃CSK₄ or PGN and supernatants assayed for TNF- α concentration. **b**, Defective Dectin-1/Syk-dependent TLR/MyD88-independent IL-10 production in *Card9*^{-/-} BMDCs. Wild-type BMDCs (WT), *Card9*^{-/-} BMDCs (-/-), wild-type BMDCs pre-treated with the Syk inhibitor piceatannol (Syk-inh.) or *Myd88*^{-/-} BMDCs (*Myd88*^{-/-}) were stimulated with zymosan and 24 h later assayed for IL-10 production. **c**, Normal IL-10 production in BMDCs from *Card9*^{+/-} and *Card9*^{-/-} mice that were left unstimulated (Medium) or stimulated with TLR ligands LPS (100 ng ml⁻¹) or CpG (1 μ M) for 24 h. The indicated values in **a–c** are means $\pm$ s.d. of triplicates and representative of three independent experiments. **d**, Zymosan-induced TNF- α , IL-6, IL-2 and IL12p40 production in wild-type BMDCs (WT), wild-type BMDCs pre-treated with Syk inhibitor (Syk-inh.) or *Card9*^{-/-} BMDCs (-/-) stimulated for 24 h as in **b**. The indicated values are



means $\pm$ s.d. of triplicates. **e**, Regular Syk activation in *Card9*^{-/-} BMDCs. Cells were stimulated with zymosan for 45 min. Syk activation was determined by immunoblot with anti-phospho-Syk antibodies. **f**, Normal Syk-dependent zymosan phagocytosis. Internalization of FITC-labelled zymosan in control (+/-) and *Card9*^{-/-} BMDCs that were pre-incubated with or without Syk inhibitor (Syk-inh.) was quantified by FACS. Percentages of cells containing FITC-labelled zymosan before (control) and after incubation are indicated. **g**, **h**, Specific defect in zymosan-induced NF- κ B activation in *Card9*^{-/-} cells. BMDCs from *Card9*^{+/-} or *Card9*^{-/-} mice were stimulated with zymosan or LPS. RelA (red) translocation into the nucleus (DAPI stained, blue; co-localization, pink) was monitored by immunofluorescence and quantified by determining the frequency of RelA-positive nuclei in at least 100 individual cells (**g**) or NF- κ B activation was determined by EMSA in BMDCs left unstimulated (medium, MED) or stimulated with zymosan (top, **h**), PGN, LPS or R848 (bottom, **h**).

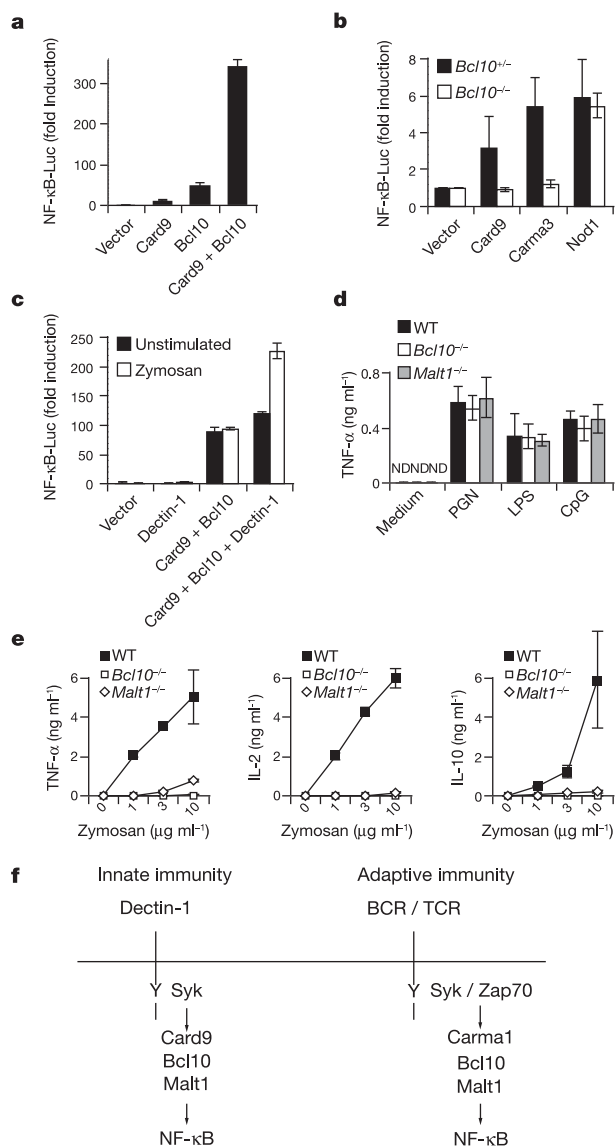


Figure 5 | Card9 signals via Bcl10 and Malt1. **a**, Card9 and Bcl10 synergize for NF- κ B activation. NF- κ B luciferase assays in 293T cells upon transfection with Card9 and/or Bcl10 expression vectors (50 ng each per well in 96 wells). Data are means $\pm$ s.d. of triplicates. **b**, Bcl10 is required for Card9-induced NF- κ B activation. Card9, Carma3 or Nod1 expression plasmids were introduced into *Bcl10*^{-/-} or *Bcl10*^{+/+} murine embryonic fibroblasts and NF- κ B luciferase assays performed. Data are means $\pm$ s.d. of triplicates. **c**, Dectin-1 ligation with zymosan augments Card9-Bcl10-induced NF- κ B induction. 293T cells were transfected with Dectin-1 (50 ng per well in 96 wells) and/or a combination of Card9 and Bcl10 expression vectors (5 ng each per well in 96 wells). NF- κ B luciferase assays were performed with or without zymosan stimulation. One representative result from four independent experiments ($\pm$ s.d.) is shown. **d**, TNF- α production in BMDCs from wild type, *Bcl10*^{-/-} or *Malt1*^{-/-} mice that were left unstimulated (Medium) or stimulated with TLR ligands PGN (20 μ g ml⁻¹), LPS (100 ng ml⁻¹) or CpG (1 μ M). ND, not detected. Data are means $\pm$ s.d. of triplicates. **e**, Bcl10 and Malt1 are required for zymosan-induced cytokine production. BMDCs from wild type, *Bcl10*^{-/-} or *Malt1*^{-/-} mice were stimulated with zymosan and TNF- α , IL-2 and IL-10 production was assayed. Results are means $\pm$ s.d. of triplicates and representative of three independent experiments. **f**, Model for the differential control of Bcl10-Malt1 signalling in innate and adaptive immunity. Y, ITAM. (See text for details.)

motif to initiate signalling via Syk through novel MyD88-independent pathways^{17,19}. The results presented here provide mechanistic insights into these pathways that can cooperate with TLR signalling for full innate immune cell activation and for the induction of pathogen-specific responses. Our findings also indicate that Bcl10-Malt1 signalling can be engaged by an ITAM-containing PRR through the adaptor Card9, presumably in an analogous manner to ITAM-containing lymphocyte antigen receptors that use Carma1 to couple to Bcl10 and Malt1. This hypothesis suggests an evolutionarily conserved mechanism for immune cell signalling and puts the concept forward that ITAM receptors in innate and adaptive immune cells can differentially connect to the conserved Bcl10-Malt1 complex via distinct CARD-coiled-coil adaptor proteins.

METHODS

Detailed descriptions of the methods are provided in Supplementary Information. **Mice, immunizations and infections.** Card9-deficient mice were generated using standard embryonic stem cell technology²⁴. To challenge the immune system, animals were either immunized intraperitoneally with TNP-Ficoll or with NP-15-Ova (Biosearch Technologies) absorbed on alum (Sigma)²⁷ or infected intravenously with *C. albicans* (strain SC5314)¹⁵ or *S. aureus* (ATCC 29219)²² as described.

Lymphocyte, BMDC and BMDM culture and cell stimulation. CD4⁺ T cells and B220⁺ B cells were purified using magnetic beads and cultured in standard lymphocyte medium²⁴. BMDCs and BMDMs were derived from bone marrow cells as described^{29,30}. Lymphocytes were stimulated with PMA and Iono (Sigma), anti-CD3, anti-CD28, anti-IgM and anti-CD40 (Becton Dickinson). BMDCs and BMDMs were stimulated with R848 (gift of S. Bauer) or Pam₃CSK₄, PGN, LPS, CpG, MDP (all from Invivogen), zymosan (from Sigma or Invivogen) or *C. albicans* strain SC5314. Cytokine concentrations in the culture supernatants were quantified using ELISA (Opt-ELIA, Becton Dickinson).

Signal transduction. Western blots using antibodies against Syk, phospho-Syk (NEB), or β -actin (Sigma), electrophoretic mobility shift assays (EMSAs) with NF- κ B-binding-site-containing oligonucleotides and NF- κ B luciferase reporter assays were performed as described^{25,27}. For immunofluorescence, BMDCs were stimulated, washed, fixed, permeabilized and incubated with anti-RelA antibody (Santa Cruz) and Alexa Fluor 594-conjugated secondary antibody, and subsequently analysed using confocal microscopy.

Received 2 April; accepted 17 May 2006.

Published online 12 July 2006.

- Akira, S. & Takeda, K. Toll-like receptor signalling. *Nature Rev. Immunol.* **4**, 499–511 (2004).
- Gordon, S. Pattern recognition receptors: doubling up for the innate immune response. *Cell* **111**, 927–930 (2002).
- Inohara, N. & Nunez, G. NODs: intracellular proteins involved in inflammation and apoptosis. *Nature Rev. Immunol.* **3**, 371–382 (2003).
- Lin, X. & Wang, D. The roles of CARMA1, Bcl10, and MALT1 in antigen receptor signaling. *Semin. Immunol.* **16**, 429–435 (2004).
- Bonizzi, G. & Karin, M. The two NF- κ B activation pathways and their role in innate and adaptive immunity. *Trends Immunol.* **25**, 280–288 (2004).
- Thome, M. CARMA1, BCL-10 and MALT1 in lymphocyte development and activation. *Nature Rev. Immunol.* **4**, 348–359 (2004).
- Bertin, J. *et al.* CARD9 is a novel caspase recruitment domain-containing protein that interacts with BCL10/CLAP and activates NF- κ B. *J. Biol. Chem.* **275**, 41082–41086 (2000).
- Matsumoto, R. *et al.* Phosphorylation of CARMA1 plays a critical role in T Cell receptor-mediated NF- κ B activation. *Immunity* **23**, 575–585 (2005).
- Sommer, K. *et al.* Phosphorylation of the CARMA1 linker controls NF- κ B activation. *Immunity* **23**, 561–574 (2005).
- Thome, M. & Tschopp, J. TCR-induced NF- κ B activation: a crucial role for Carma1, Bcl10 and MALT1. *Trends Immunol.* **24**, 419–424 (2003).
- Di Carlo, F. J. & Fiore, J. V. On the composition of zymosan. *Science* **127**, 756–757 (1958).
- Brown, G. D. & Gordon, S. Immune recognition. A new receptor for β -glucans. *Nature* **413**, 36–37 (2001).
- Underhill, D. M. *et al.* The Toll-like receptor 2 is recruited to macrophage phagosomes and discriminates between pathogens. *Nature* **401**, 811–815 (1999).
- Brown, G. D. & Gordon, S. Fungal β -glucans and mammalian immunity. *Immunity* **19**, 311–315 (2003).
- Villamon, E. *et al.* Toll-like receptor-2 is essential in murine defenses against *Candida albicans* infections. *Microbes Infect.* **6**, 1–7 (2004).
- Gantner, B. N., Simmons, R. M., Canavera, S. J., Akira, S. & Underhill, D. M. Collaborative induction of inflammatory responses by dectin-1 and Toll-like receptor 2. *J. Exp. Med.* **197**, 1107–1117 (2003).

17. Rogers, N. C. *et al.* Syk-dependent cytokine induction by Dectin-1 reveals a novel pattern recognition pathway for C type lectins. *Immunity* **22**, 507–517 (2005).
18. Brown, G. D. *et al.* Dectin-1 mediates the biological effects of β -glucans. *J. Exp. Med.* **197**, 1119–1124 (2003).
19. Brown, G. D. Dectin-1: a signalling non-TLR pattern-recognition receptor. *Nature Rev. Immunol.* **6**, 33–43 (2005).
20. Takeuchi, O. *et al.* Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. *Immunity* **11**, 443–451 (1999).
21. Aliprantis, A. O. *et al.* Cell activation and apoptosis by bacterial lipoproteins through toll-like receptor-2. *Science* **285**, 736–739 (1999).
22. Takeuchi, O., Hoshino, K. & Akira, S. Cutting edge: TLR2-deficient and MyD88-deficient mice are highly susceptible to *Staphylococcus aureus* infection. *J. Immunol.* **165**, 5392–5396 (2000).
23. Kawai, T., Adachi, O., Ogawa, T., Takeda, K. & Akira, S. Unresponsiveness of MyD88-deficient mice to endotoxin. *Immunity* **11**, 115–122 (1999).
24. Ruland, J. *et al.* Bcl10 is a positive regulator of antigen receptor-induced activation of NF- κ B and neural tube closure. *Cell* **104**, 33–42 (2001).
25. McAllister-Lucas, L. M. *et al.* Bimp1, a MAGUK family member linking protein kinase C activation to Bcl10-mediated NF- κ B induction. *J. Biol. Chem.* **276**, 30589–30597 (2001).
26. Eliopoulos, A. G., Das, S. & Tsichlis, P. N. The tyrosine kinase Syk regulates TPL2 activation signals. *J. Biol. Chem.* **281**, 1371–1380 (2006).
27. Ruland, J., Duncan, G. S., Wakeham, A. & Mak, T. W. Differential requirement for Malt1 in T and B cell antigen receptor signaling. *Immunity* **19**, 749–758 (2003).
28. Klemm, S. *et al.* The Bcl10-Malt1 complex segregates Fc ϵ RI-mediated nuclear factor κ B activation and cytokine production from mast cell degranulation. *J. Exp. Med.* **203**, 337–347 (2006).
29. Hammer, M. *et al.* Control of dual-specificity phosphatase-1 expression in activated macrophages by IL-10. *Eur. J. Immunol.* **35**, 2991–3001 (2005).
30. Sparwasser, T. *et al.* Bacterial DNA and immunostimulatory CpG oligonucleotides trigger maturation and activation of murine dendritic cells. *Eur. J. Immunol.* **28**, 2045–2054 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Wagner, R. Lang, R. Rupec and M. Thome for discussions; S. Bauer and K. Pechloff for critically reading the manuscript; B. Holzmann for providing Myd88^{-/-} bone marrow; A. Walch for access to a confocal microscope; M. Neuenhahn for help with intravenous injections; and S. Weiss, S. Leeder and K. Meiners for technical assistance. This work was supported by SFB grants from Deutsche Forschungsgemeinschaft to I.F. and J.R. and by a Max-Eder-Program grant from Deutsche Krebshilfe to J.R.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.R. (jruland@lrz.tum.de).

A probable stellar solution to the cosmological lithium discrepancy

A. J. Korn¹, F. Grundahl², O. Richard³, P. S. Barklem¹, L. Mashonkina⁴, R. Collet¹, N. Piskunov¹ & B. Gustafsson¹

The measurement of the cosmic microwave background has strongly constrained the cosmological parameters of the Universe¹. When the measured density of baryons (ordinary matter) is combined with standard Big Bang nucleosynthesis calculations^{2,3}, the amounts of hydrogen, helium and lithium produced shortly after the Big Bang can be predicted with unprecedented precision^{1,4}. The predicted primordial lithium abundance is a factor of two to three higher than the value measured in the atmospheres of old stars^{5,6}. With estimated errors of 10 to 25%, this cosmological lithium discrepancy seriously challenges our understanding of stellar physics, Big Bang nucleosynthesis or both. Certain modifications to nucleosynthesis have been proposed⁷, but found experimentally not to be viable⁸. Diffusion theory, however, predicts atmospheric abundances of stars to vary with time⁹, which offers a possible explanation of the discrepancy. Here we report spectroscopic observations of stars in the metal-poor globular cluster NGC 6397 that reveal trends of atmospheric abundance with evolutionary stage for various elements. These element-specific trends are reproduced by stellar-evolution models with diffusion and turbulent mixing¹⁰. We thus conclude that diffusion is predominantly responsible for the low apparent stellar lithium abundance in the atmospheres of old stars by transporting the lithium deep into the star.

Diffusive processes altering the elemental composition in stars have been studied for decades^{9,11}. Evidence for their importance comes from helioseismology¹² and the study of hot stars with peculiar abundance patterns¹³. Among solar-type stars, the effects of diffusion are expected to be more pronounced in old, very metal-poor stars. Given their greater age, diffusion has had more time to produce sizeable effects than in younger stars like the Sun. Detailed element-by-element predictions from models including effects of atomic diffusion and radiative accelerations became available a few years ago¹⁴, but these early models produced strong abundance trends that are incompatible with measurements of, in particular, the abundance of lithium common among stars of the Galactic halo over a wide range of metallicities (the so-called Spite plateau of lithium). However, the recent inclusion of turbulent mixing¹⁰ brings model predictions into better agreement with observations.

According to the predictions from such model calculations, stars leaving the main-sequence (turn-off stars) are expected to show the largest variations relative to the composition of the gas from which the stars originated. Giant stars, however, have deep surface convection zones which erase most effects of diffusion and restore the original composition. One notable exception is lithium which disintegrates in layers with $T > 2.1$ million K. The destruction of lithium inside the star leads to a successive dilution of the surface lithium because the convective envelope expands when the star

becomes a red giant. We performed spectroscopic observations specifically to test these model predictions.

Globular-cluster stars have identical age and initial heavy-element composition, and thus measured atmospheric abundance trends with evolutionary stage are a signature of diffusion. We observed 18 stars along the evolutionary sequence of NGC 6397 with the multi-object spectrograph FLAMES-UVES on the 8.2-m Very Large Telescope in Chile—five stars close to the main-sequence turn-off point (TOP), two stars in the middle of the subgiant branch (SGB), five stars at the base of the red-giant branch (bRGB) and six red giants (RGB)—that represent specific stages in stellar evolution (Supplementary Fig. 1). Even on 8-m-class telescopes, these observations are still challenging, requiring total integration times of 2–12 hours per star to obtain homogeneous data with high signal-to-noise ratios at high spectral resolution.

These observational challenges explain in part why there have been surprisingly few studies looking for abundance trends between unevolved and evolved stars in globular clusters. One study¹⁵ found lower abundances for subgiants in the very metal-poor globular cluster M92 than accepted literature values indicate for giant stars. However, the low signal-to-noise data did not allow firm conclusions. A different study¹⁶ examined TOP stars and subgiant stars in NGC 6397. It concentrated on elements showing suspected nuclear processing (oxygen, sodium, magnesium and aluminium). The analysis was based on standard models of stellar evolution and spectroscopic analysis (assuming equilibrium) and did not indicate significant differences in iron abundances. This result critically depends on the above-mentioned assumptions, which yield stellar parameters (effective temperatures) unsupported by photometry. The only previous study attempting a homogeneous analysis of red-giant and dwarf stars¹⁷ targeted M13, a globular cluster a factor of three more metal-rich than NGC 6397. The highest-gravity object (a subgiant with $\log g = 3.8$) indeed shows an iron abundance 41% (0.23 in log abundance) below the average of the other 24 more-evolved cluster members analysed. This was not interpreted as an indication of departures from a uniform cluster composition as caused by diffusion.

To derive stellar parameters and elemental abundances from our observations, we employed well-established spectroscopic diagnostics (in one dimension) with a high level of modelling realism in the line formation (non-equilibrium where necessary). The profile of the Balmer line of hydrogen ($H\alpha$, at 656 nm) was used to estimate the effective temperatures (T_{eff})^{18,19}, while the ionization equilibrium of iron (treated in non-equilibrium) was used to determine the stellar surface gravities ($\log g$)²⁰. The mean stellar parameters for the four groups of stars are given in Table 1.

An independent check of the reliability of the stellar-parameter

¹Department of Astronomy and Space Physics, Uppsala University, Box 515, 75120 Uppsala, Sweden. ²Department of Physics and Astronomy, University of Århus, Ny Munkegade, 8000 Århus C, Denmark. ³GRAAL-UMR5024/ISTEEM (CNRS), Université Montpellier II, Place E. Bataillon, 34095 Montpellier, France. ⁴Institute of Astronomy, Russian Academy of Science, Pyatnitskaya 48, 119017 Moscow, Russia.

determination is obtained based on photometry, exploiting the luminosity difference of the stars. These data are also given in Table 1. As can be seen, the spectroscopic $\Delta T_{\text{eff}}(\text{TOP-RGB})$ (that is, $T_{\text{eff, TOP}} - T_{\text{eff, RGB}}$) is well-matched by both the Strömgren index ($v - y$) and the broad-band index ($V - I$), the photometry indicating a ΔT_{eff} only 20–50 K (2–5%) lower. The spectroscopic $\Delta \log g(\text{TOP-RGB})$ is only 0.05 below that determined from the photometry. This is excellent agreement for two fully independent methods and we thus consider the stellar-parameter differences to be well constrained. On the basis of these differences, relative abundance trends can be scrutinized.

The best-determined abundance (in terms of number of spectral lines used) is that of iron. We find the abundance to differ by 45% (0.16 in log abundance), that is, the TOP stars have the lowest abundance which successively rises towards the RGB stars (Fig. 1a). Propagating the line-to-line scatter for all iron lines of each star into the mean values given in Table 1 and further into the abundance difference between TOP and RGB stars, the difference ($\Delta \log \epsilon_{\text{TOP-RGB}} = 0.16 \pm 0.05$) is significant at the 3.2σ level. Calcium and titanium also show trends (Supplementary Fig. 2), but these are less pronounced.

We have compared these abundance trends with various diffusion model predictions and found that a model with one particular value of the turbulent-mixing efficiency¹⁰ is capable of fitting the observations of these heavy elements well. This model (model T6.0, with a parametric description of turbulent mixing using an isotropic turbulent diffusion coefficient 400 times larger than the atomic diffusion coefficient for helium at $\log T = 6.0$ varying with density as ρ^{-3} (ref. 10), and computed for the metallicity of the RGB stars) clearly predicts a steeper trend for iron than for calcium or titanium, in agreement with the observations. The simultaneous element-specific reproduction of the observations thus lends strong support for the diffusion interpretation of these trends and constrains the level of turbulent mixing.

We emphasize that it is not possible to remove all trends simultaneously by adjusting the stellar parameters: the neutral species Ca I and Fe I (both affected by non-equilibrium) require a T_{eff} correction of roughly 100 K and 200 K respectively, while the ionized species Ti II and Fe II (formed in equilibrium) would require a change in $\log g$ of 0.15 and 0.33, respectively. Considering the agreement between

spectroscopy and photometry, these corrections to the stellar parameters are large and incompatible with one another.

We also investigate the effect that the three-dimensional (3D) hydrodynamic nature of stellar atmospheres²¹ might have on the observed trends. Using the 3D TOP and RGB models available to us, we find that the trends inferred from weak Fe II and Ti II lines in one dimension require small 3D corrections only. For iron, the trend would even be slightly steeper. As the stellar-parameter differences are well determined, analyses based on 3D hydrodynamic models would reach much the same conclusion about the abundance trends.

In the light of the identified diffusion signature, we should consider the structural effect that helium diffusion has on the atmosphere and, therefore, on the spectroscopic analysis. The T6.0 model predicts the He/H ratio in the TOP stars to be decreased by nearly 50% from the original value, 40% in the SGB stars, whereas the original ratio is almost restored in the bRGB and RGB stars. Helium settling of this extent changes the mean molecular weight in the

Table 1 | Mean stellar parameters, iron abundances and photometric quantities of the four groups of stars observed

Group	No. of stars	T_{eff} (K)	$\log[g \text{ (cm s}^{-2}\text{)}]$	$\log[\epsilon(\text{Fe})]$	ξ (km s ⁻¹)
TOP	5	6,254	3.89	5.23 ± 0.04	2.00
SGB	2	5,805	3.58	5.27 ± 0.05	1.75
bRGB	5	5,456	3.37	5.33 ± 0.03	1.73
RGB	6	5,130	2.56	5.39 ± 0.02	1.60
Sun	1	5,777	4.44	7.51	1.00

	$\Delta T_{\text{eff}}(\text{TOP-RGB})$ (K)	$\Delta \log[g(\text{TOP-RGB})]$ (cm s ⁻²)	$\Delta \log[\epsilon(\text{Fe})_{(\text{TOP-RGB})}]$
Spectroscopy	1,124	1.33	0.16 ± 0.05
Strömgren, (v-y)	1,108	1.38	-
Broad-band, (V-I)	1,070	1.38	-

Mean stellar parameters of the four groups of stars in the following evolutionary stages: TOP, SGB, bRGB and RGB (see text). The FLAMES-UVES spectra cover the spectral range from 4,800–6,800 Å, have $R = \lambda/\Delta\lambda = 48,000$ (where λ is wavelength) and signal-to-noise ratios in excess of 80:1 per pixel. The analyses are fully spectroscopic and line-by-line differential to the Sun. Typical errors on T_{eff} , $\log g$ and the microturbulence ξ for individual stars are, respectively, 150 K, 0.15 and 0.2 km s⁻¹. The errors in $\log[\epsilon(\text{Fe})] = \log(N_{\text{Fe}}/N_{\text{H}}) + 12$ are the combined values of the line-to-line scatter of Fe I and Fe II for the individual stars propagated into the mean value for the group. Between 20 and 40 Fe I and Fe II lines were measured by means of profile fits. The stellar parameters of the TOP and SGB group have not been corrected for helium diffusion, which would result in higher $\log g$ values (+0.05; see text). Below, the spectroscopic results are compared with the photometry (Strömgren and broad-band indices calibrated on the infrared-flux method²⁸, see also Supplementary Table 1) obtained with the Danish 1.54-m telescope on La Silla. In both cases, $\Delta \log g$ is based on the magnitude difference in V.

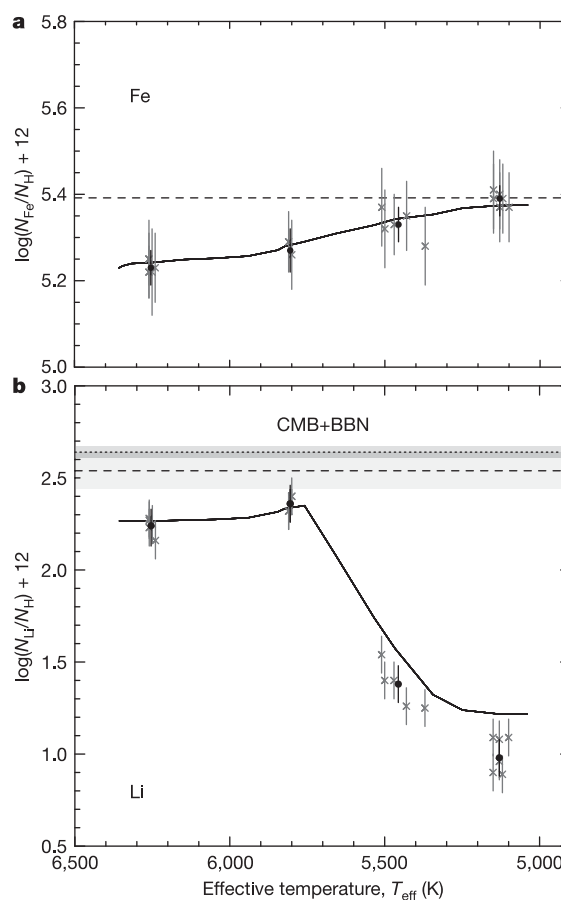


Figure 1 | Trends of iron and lithium as a function of the effective temperatures of the observed stars compared to the model predictions. The grey crosses are the individual measurements, while the bullets are the group averages. The solid lines are the predictions of the diffusion model, with the original abundance given by the dashed line. In **b**, the grey-shaded area around the dotted line indicates the 1σ confidence interval of CMB + BBN¹: $\log[\epsilon(\text{Li})] = \log(N_{\text{Li}}/N_{\text{H}}) + 12 = 2.64 \pm 0.03$. In **a**, iron is treated in non-equilibrium²⁰ (non-LTE), while in **b**, the equilibrium (LTE) lithium abundances are plotted, because the combined effect of 3D and non-LTE corrections was found to be very small²⁹. For iron, the error bars are the line-to-line scatter of Fe I and Fe II (propagated into the mean for the group averages), whereas for the absolute lithium abundances 0.10 is adopted. The 1σ confidence interval around the inferred primordial lithium abundance ($\log[\epsilon(\text{Li})] = 2.54 \pm 0.10$) is indicated by the light-grey area. We attribute the modelling shortcomings with respect to lithium in the bRGB and RGB stars to the known need for extra mixing³⁰, which is not considered in the diffusion model.

atmosphere appreciably, an effect which can be mapped as a shift in gravity²². The accompanying shift in $\log g$ amounts to +0.05 for both groups, bringing the spectroscopic $\Delta \log g_{\text{TOP-RGB}}$ into perfect agreement with the photometry-based value. This correction also improves the agreement of the stellar parameters with the isochrone (Supplementary Fig. 1).

Lithium is ionized easily and therefore behaves much like helium in terms of settling. As seen from Fig. 1b, both the TOP stars and the SGB stars sample the lithium plateau. The observations support the T6.0 model predictions, with a slight upturn towards cooler temperatures before the convection zone encompasses lithium-free layers and dilution sets in. A higher value of lithium among halo subgiants than among halo dwarf stars was recently reported²³ and qualitatively interpreted as a signature of atomic diffusion counterbalanced by gravity-wave-induced mixing. In quantitative terms, the diffusion model we use predicts the original lithium abundance to be 78% (0.25 in log abundance) higher than the mean value measured in the TOP and SGB stars. Therefore, we infer a primordial lithium abundance which agrees with the cosmic microwave background plus Big Bang nucleosynthesis (CMB + BBN) value¹ within the mutual 1σ error bars. Given the uncertainties associated with both values, the cosmological lithium discrepancy is thus largely removed. This finding restores confidence in standard BBN, quantitative spectroscopy and sophisticated stellar evolution models.

In extrapolating to the primordial lithium abundance, we do not consider corrections related to the Galactic chemical evolution of lithium. Such corrections are model-dependent and rather uncertain. It is plausible that some fraction of the halo gas was processed through Population III stars²⁴ (lowering the lithium abundance) and that cosmic rays interacting with the interstellar medium have produced some lithium⁶. At present it is, however, unclear to which extent (and even in which direction) the stellar lithium abundances should be adjusted.

While helium diffusion has been invoked to lower globular-cluster ages to values which meet the cosmological age constraint, metal diffusion has rarely been considered. The models including atomic diffusion, radiative accelerations and turbulent mixing for all elements will result in moderately different absolute and relative ages. Using similar models²⁵, the analysis of M92 has led to an age estimate compatible with Wilkinson Microwave Anisotropy Probe (WMAP¹) results. It is clear, however, that stellar-evolution models need to be developed further towards physical self-consistency with respect to all relevant processes, including hydrodynamics and rotation. In particular, the physical mechanism causing turbulent mixing should be identified. Promising work²⁶ points towards the importance of internal gravity waves in transporting angular momentum and mixing stellar interiors.

Regarding unevolved metal-poor stars as tracers of Galactic cosmochemistry, heavy-element abundance ratios (for example, Mg/Fe) are less affected by diffusion than ratios with respect to hydrogen (for example, Fe/H). However, the effects are non-negligible and must be taken into account to reach the highest accuracy. Care should be taken especially when comparing abundances from giant and dwarf stars. In this respect, one exceptional pair is the two most metal-poor stars known²⁷, whose abundance signatures will have to be reinvestigated in the light of diffusion.

Received 28 March; accepted 14 June 2006.

1. Spergel, D. N. *et al.* Wilkinson Microwave Anisotropy Probe (WMAP) three year results: implications for cosmology. *Astrophys. J.* (submitted); preprint at (<http://arxiv.org/abs/astro-ph/0603449>) (2006).
2. Wagoner, R. V., Fowler, W. A. & Hoyle, F. On the synthesis of elements at very high temperatures. *Astrophys. J.* **148**, 3–49 (1967).
3. Burles, S., Nollett, K. M. & Turner, M. S. Big bang nucleosynthesis predictions for precision cosmology. *Astrophys. J.* **552**, L1–L5 (2001).
4. Cyburt, R. H., Fields, B. D. & Olive, K. A. Primordial nucleosynthesis in light of WMAP. *Phys. Lett. B* **567**, 227–234 (2003).
5. Spite, M. & Spite, F. Lithium abundance at the formation of the galaxy. *Nature* **297**, 483–485 (1982).
6. Ryan, S. G., Norris, J. E. & Beers, T. C. The Spite lithium plateau: ultrathin but postprimordial. *Astrophys. J.* **523**, 654–677 (1999).
7. Coc, A., Vangioni-Flam, E., Descouvemont, P., Adahchour, A. & Angulo, C. Updated big bang nucleosynthesis compared with Wilkinson Microwave Anisotropy Probe observations and the abundance of light elements. *Astrophys. J.* **600**, 544–552 (2004).
8. Angulo, C. *et al.* The ${}^7\text{Be}(d,p){}^2\alpha$ cross section at big bang energies and the primordial ${}^7\text{Li}$ abundance. *Astrophys. J.* **630**, L105–L108 (2005).
9. Aller, L. H. & Chapman, S. Diffusion in the sun. *Astrophys. J.* **132**, 461–472 (1960).
10. Richard, O., Michaud, G. & Richer, J. Implications of WMAP observations on Li abundance and stellar evolution models. *Astrophys. J.* **619**, 538–548 (2005).
11. Michaud, G., Fontaine, G. & Beaudet, G. The lithium abundance: constraints on stellar evolution. *Astrophys. J.* **282**, 206–213 (1984).
12. Guzik, J. A. & Cox, A. N. On the sensitivity of high-degree p-mode frequencies to the solar convection zone helium abundance. *Astrophys. J.* **386**, 729–733 (1992).
13. Richer, J., Michaud, G. & Turcotte, S. The evolution of AMFM stars, abundance anomalies, and turbulent transport. *Astrophys. J.* **529**, 338–356 (2000).
14. Richard, O., Michaud, G. & Richer, J. Models of metal-poor stars with gravitational settling and radiative accelerations. III. Metallicity dependence. *Astrophys. J.* **580**, 1100–1117 (2002).
15. King, J. R., Stephens, A., Boesgaard, A. M. & Deliyannis, C. Keck HIRES spectroscopy of M92 subgiants—surprising abundances near the turnoff. *Astron. J.* **115**, 666–684 (1998).
16. Gratton, R. G. *et al.* The O-Na and Mg-Al anticorrelations in turn-off and early subgiants in globular clusters. *Astron. Astrophys.* **369**, 87–98 (2001).
17. Cohen, J. G. & Meléndez, J. Abundances in a large sample of stars in M3 and M13. *Astron. J.* **129**, 303–329 (2005).
18. Fuhrmann, K., Axer, M. & Gehren, T. Balmer lines in cool dwarf stars. I. Basic influence of atmospheric models. *Astron. Astrophys.* **271**, 451–462 (1993).
19. Barklem, P. S., Piskunov, N. & O'Mara, B. J. Self-broadening in Balmer line wing formation in stellar atmospheres. *Astron. Astrophys.* **363**, 1091–1105 (2000).
20. Korn, A. J., Shi, J. & Gehren, T. Kinetic equilibrium of iron in the atmospheres of cool stars. III. The ionization equilibrium of selected reference stars. *Astron. Astrophys.* **407**, 691–703 (2003).
21. Stein, R. F. & Nordlund, Å. Simulations of solar granulation. I. General properties. *Astrophys. J.* **499**, 914–933 (1998).
22. Strömgren, B., Gustafsson, B. & Olsen, E. H. Evidence of helium abundance differences between the Hyades stars and field stars, and between Hyades stars and Coma cluster stars. *Astron. Soc. Pacif.* **94**, 5–15 (1982).
23. Charbonnel, C. & Primas, F. The lithium content of the galactic halo stars. *Astron. Astrophys.* **442**, 961–992 (2005).
24. Piau, L. *et al.* From first stars to the Spite plateau: a possible reconciliation of halo stars observations with predictions from big bang nucleosynthesis. *Astrophys. J.* (submitted); preprint at (<http://arxiv.org/abs/astro-ph/0603553>) (2006).
25. Vandenberg, D. A., Richard, O., Michaud, G. & Richer, J. Models of metal-poor stars with gravitational settling and radiative accelerations. II. The age of the oldest stars. *Astrophys. J.* **571**, 487–500 (2002).
26. Charbonnel, C. & Talon, S. Influence of gravity waves on the internal rotation and Li abundance of solar-type stars. *Science* **309**, 2189–2191 (2005).
27. Frebel, A. *et al.* Nucleosynthetic signatures of the first stars. *Nature* **434**, 871–873 (2005).
28. Alonso, A., Arribas, S. & Martinez-Roger, C. The effective temperature scale of giant stars (F0–K5). II. Empirical calibration of T_{eff} versus colours and [Fe/H]. *Astron. Astrophys. Suppl.* **140**, 261–277 (1999).
29. Barklem, P. S., Belyaev, A. K. & Asplund, M. Inelastic H + Li and $\text{H}^- + \text{Li}^+$ collisions and non-LTE Li I line formation in stellar atmospheres. *Astron. Astrophys.* **409**, L1–L4 (2003).
30. Charbonnel, C. A consistent explanation for ${}^{12}\text{C}/{}^{13}\text{C}$, ${}^7\text{Li}$ and ${}^3\text{He}$ anomalies in red giant stars. *Astrophys. J.* **453**, L41–L44 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A.J.K. acknowledges a research fellowship by the Leopoldina Foundation, Germany. O.R. thanks the Centre Informatique National de l'Enseignement Supérieur (CINES) and the Réseau Québécois de Calcul de Haute Performance (RQCHP) for providing the computational resources required for this work. F.G. acknowledges financial support from the Instrument Center for Danish Astrophysics (IDA). L.M. acknowledges support through the Presidium RAS Programme 'Origin and evolution of stars and the Galaxy'. The Uppsala group of authors acknowledges support from the Swedish Research Council. We thank A. Alonso and I. Ramirez for providing colour-temperature relations specific to this project.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.J.K. (akorn@astro.uu.se).

LETTERS

Occultation of X-rays from Scorpius X-1 by small trans-neptunian objects

Hsiang-Kuang Chang^{1,2}, Sun-Kun King³, Jau-Shian Liang¹, Ping-Shien Wu², Lupin Chun-Che Lin¹
& Jeng-Lun Chiu¹

Since the discovery¹ of the trans-neptunian objects (TNOs) in 1992, nearly one thousand new members have been added to our Solar System^{2,3}, several of which are as big as—or even larger than—Pluto^{4,5}. The properties of the population of TNOs, such as the size distribution and the total number, are valuable information for understanding the formation of the Solar System, but direct observation is only possible for larger objects with diameters above several tens of kilometres. Smaller objects, which are expected to be more abundant, might be found when they occult background stars^{6–10}, but hitherto there have been no definite detections. Here we report the discovery of such occultation events at millisecond timescales in the X-ray light curve of Scorpius X-1. The estimated sizes of these occulting TNOs are ≤ 100 m. Their abundance is in line with an extrapolation of the distribution² of sizes of larger TNOs.

Because of the photon-counting nature of X-ray observations, X-ray occultation may stand a good chance of revealing the existence of small TNOs as long as the background X-ray source is bright enough to allow statistically meaningful determination at short timescales. Scorpius X-1 is the brightest and first-discovered X-ray source outside our Solar System¹¹. It has been extensively studied and observed, particularly by the Rossi X-ray Timing Explorer (RXTE). In addition, Sco X-1 is only about 6° to the north of the ecliptic, making it the best target for a TNO occultation search. RXTE is a NASA mission launched at the end of 1995, providing the largest effective area of all X-ray satellites¹².

The Proportional Counter Array (PCA) instrument (2–60 keV) onboard RXTE registers a raw count rate of about 10^5 counts per second from Sco X-1. Although PCA, with a one-degree field of view, has no imaging capability, the extremely high count rate from the point source Sco X-1 makes the sky and instrumental background both negligible. This high count rate also enables us to perform an examination on its light curve at timescales as short as one millisecond to search for possible occultation.

To search for occultation events, we examine all the time bins in the available RXTE/PCA archival data of Sco X-1 and compute their deviation distribution (Fig. 1). The deviation distribution of all the time bins in the light curve should be poissonian, modified by the dead-time and coincidence-event effects if all the fluctuations are random and there is no occultation event. The PCA data we used span over seven years from 1996 to 2002 and the total exposure time is about 322 ks. A log of all these data are listed in Supplementary Table 1.

In our search, we found obvious excess at the negative-deviation end in the deviation distribution. The excess becomes less significant for cases of larger bin sizes. Two of these distributions are plotted in Fig. 1. To check whether these negative-deviation excesses are instrumental, we performed the same search with PCA data for the

Crab nebula, which is an extended X-ray source and is thus most suitable for checking whether those excesses are instrumental rather than due to occultations. Although the PCA count rate of the Crab nebula is about one-eighth of that of Sco X-1, a similar excess should still be detectable with enough data if it is instrumental in origin. From our search on 380-ks PCA data of the Crab nebula, no such excess was found. We also searched the PCA data of a candidate

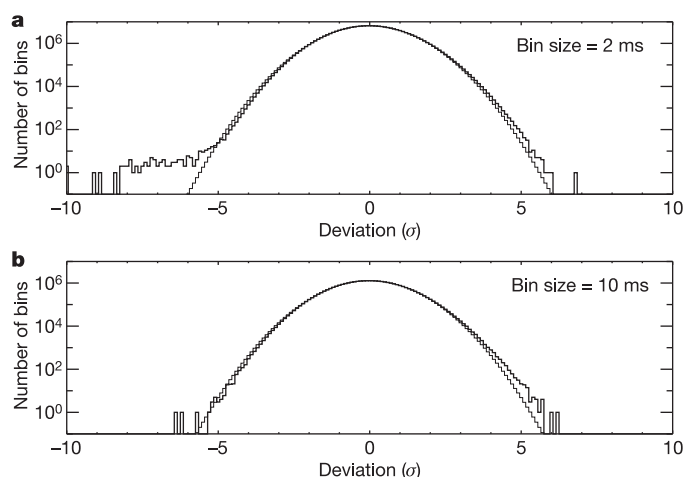


Figure 1 | Deviation distributions of the Sco X-1 RXTE/PCA light curve.

The 'deviation' is defined for each time bin in the data as the difference between the photon number counts of that time bin and the mean counts per bin in a time window encompassing and running with the time bin in question. The 'deviation' is further expressed in units of the standard deviation of the photon counts per bin in the associated running window. In practice, the size of the running window should be large enough to have enough bins for the determination of the running mean and the standard deviation. However, it should not be too large, or the intrinsic flux level change of Sco X-1 will be involved. An adequate window size can be determined from comparing simulated distributions with computed (theoretically expected) ones. Searches for large deviation have been performed with all the available RXTE archival data, using different time-bin sizes from 1 to 10 ms with a window size of 8 s and time-bin sizes from 11 to 20 ms with a window size of 16 s. Here we show the deviation distributions for bin sizes of 2 ms (**a**) and 10 ms (**b**). Thick histograms are from the 322-ks RXTE/PCA data of Sco X-1 and thin histograms are gaussian distributions, plotted for comparison. To show the relatively small number of dip events, the ordinate is plotted on logarithmic scales. The excess at the negative deviation end in **a** is obvious. It becomes insignificant when the bin size increases to 10 ms. The small excess at the positive deviation part in both panels and the tiny deficiency between about -3σ and -5σ are due to the Poisson nature of the observed photons.

¹Department of Physics, National Tsing Hua University, Hsinchu 30013, Taiwan. ²Institute of Astronomy, National Tsing Hua University, Hsinchu 30013, Taiwan. ³Institute of Astronomy and Astrophysics, Academia Sinica, Taipei 10617, Taiwan.

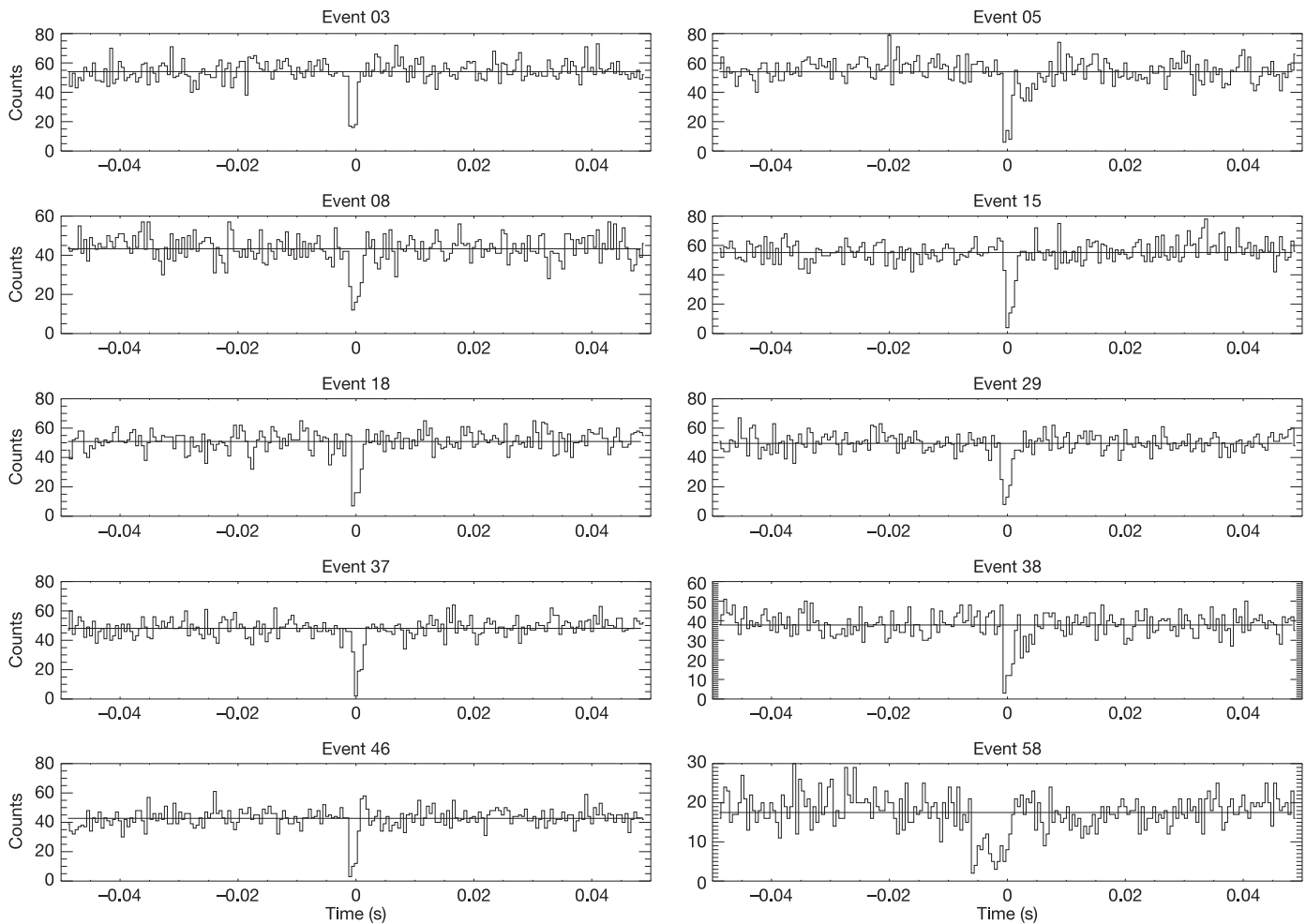


Figure 2 | RXTE/PCA light curves of some example dip events. ‘Dip events’ are identified as those time bins with a large negative deviation whose random probability is lower than 10^{-3} . Assuming a gaussian probability distribution for simplicity, it corresponds to a negative deviation of about 6.8σ in the deviation distribution, depending on the total number of all time bins in that search, which, for example, is about 1.6×10^8 for the search with 2-ms time bins in the 322-ks data. Dip events may be repeatedly identified in searches with different bin sizes. In total, 58 dip events are found in the 322-ks data. We show the light curves of ten of the most significant events. Horizontal lines are the average counts in each panel. The histograms are

accreting black hole, 4U 1543-47, which had an outburst in 1996 with a flux level about two to three times that of the Crab nebula for 4 ks. No negative-deviation excess was found for 4U 1543-47 either. Furthermore, as shown in Fig. 2, the count number drops in these time bins by a very large factor, indicating that these counts are not caused by malfunction of any single one of the five independent, almost-identical Proportional Counter Units (PCUs) of the PCA, although the recorded data mode does not contain individual PCU information. We therefore consider the excess very unlikely to be of instrumental origins.

In the 322-ks RXTE/PCA data of Sco X-1, 58 dip events with random probability lower than 10^{-3} were found. Some dip events are very significant, with deviation at a 10σ level. For example, event 18 in Fig. 2, when binned with a 2-ms bin size, has a deviation of -10.7σ . It corresponds to a random probability lower than 10^{-18} with the total number of time bins ($\sim 1.6 \times 10^8$) taken into account. The light curves of some of the most significant dip events are shown in Fig. 2. All 58 light curves can be found in Supplementary Fig. 1.

Sco X-1 is very bright. Almost all of the recorded multiple-anode events (events triggering more than two anodes in one PCU) are

plotted with a bin size of 0.5-ms. Events are numbered in sequence according to their occurrence time. Different count rates for different observations are due to the intrinsic flux variation of Sco X-1 and the number of PCUs turned on in that observation. Among the 58 events, more than a quarter show an asymmetric occultation light curve with fast drop in photon counts followed by slower recovery, like events 18 and 46 shown here. Only three are asymmetric in the opposite sense. Four events possibly show a complicated light curve like events 5 and 38 shown here. The others are roughly symmetric. All the light curves and more information on the 58 events can be found in Supplementary Fig. 1 and Supplementary Table 2.

actually caused by two or more photons from Sco X-1 arriving simultaneously (within a coincidence window, typically of $5\mu\text{s}$; ref. 13). These multiple-anode events are usually rejected. Extremely bright flares might therefore be recorded as dips. However, the dips we found are not caused by flares, because the number of recorded two-anode events, which trigger both the L1 and R1 anodes in one PCU, drops at the same time for all 58 dip events, instead of increasing dramatically.

Most events have a duration of 2–3 ms and the longest one is 7 ms. The occurrence time of these 58 events seems random, based on the currently available data. The epochs and duration of all the 58 events are listed in Supplementary Table 2. We have also examined possible colour changes for the 33 events whose spectral information is available in a two-channel manner. We compared the ratio of counts in the two spectral channels with 200-ms bins, averaged over 4 s before and after the event. No significant colour change was found in any of these events.

It is difficult to attribute these dips to the intrinsic variability of Sco X-1. The luminosity variation of Sco X-1 at timescales from milliseconds to hours has been known for a long time^{14–16}. These

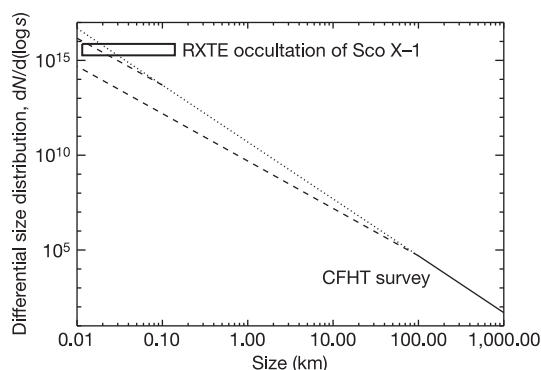


Figure 3 | The differential size distribution of TNOs. The total number of TNOs per decade of size— $dN/d(\log s)$ —inferred from the RXTE occultation of Sco X-1 is plotted as the bold rectangle at the upper left. Because of various assumptions made in this estimate (see Supplementary Discussion), an uncertainty level of a factor of two is adopted in this plot. The solid line ($dN/ds \propto s^{-q}$, where $q = 4.0$) near the lower right is the best fit of the Canada-France-Hawaii Telescope (CFHT) survey². Its extrapolation towards smaller size is plotted as a dotted line. Two dashed lines corresponding to $q = 3.5$, as suggested by collisional erosion models^{23,24}, are plotted with turnover size at 0.1 and 100 km for comparison. Direct imaging of TNOs down to about 20 km in size can be achieved by the Hubble Space Telescope²⁵. The Taiwanese-American Occultation Survey⁹ (TAOS) is expected to detect occultation events caused by TNOs of kilometre size. Our discovery of the X-ray occultation demonstrates a new method of studying even smaller TNOs.

variations happen with a small amplitude at all times in different activity states of the source, unlike the dip events we report here, which are sparse, abrupt, and always accompany a significant drop in flux. Millisecond X-ray flares have been observed in Cyg X-1 (ref. 17), and are very likely to be a manifestation of the lognormal nature of its X-ray flux¹⁸. The millisecond dips we report here are, however, dips, not flares, and the flux fluctuation of Sco X-1 is poissonian rather than lognormal.

There is strong evidence indicating that Sco X-1 contains a neutron star accreting from a companion of 0.4 solar masses¹⁹ and the observed X-ray comes from the neutron-star surface and the inner part of the accretion disk, whose variability is related to the change of the accretion rate^{20,21}. A sudden quench of plasma corona near the inner accretion disk of Sco X-1 might produce such dips but this would require an unusual cooling scheme. Objects such as neutron stars or stellar-mass black holes would need to have a transverse speed of several thousand kilometres per second to block Sco X-1 for a short duration of several milliseconds, and are thus also unlikely to cause the dips we observed.

We thus conclude that occultation by objects in the Solar System is the most plausible explanation to these dip events. The dimension of the X-ray-emitting region of Sco X-1 is controversial. The estimation ranges from 100 km to about 50,000 km, depending on different underlying models^{20,21}. With a distance of 2.8 kpc (ref. 22), Sco X-1 occupies, at the most, a size of only 3 m at 40 AU (astronomical units) and more probably occupies only a few centimetres. It is even smaller at the distance of the main-belt asteroids. Taking Sco X-1 as a point source, the estimated occultation rate by the main-belt asteroids is extremely low (less than 10^{-9} s^{-1} ; details in the Supplementary Discussion). On the other hand, numerous larger TNOs have been discovered. An estimated total number of TNOs larger than 100 km within 30–50 AU is about 4×10^4 (ref. 2). Their differential size distribution follows the power law $dN/ds \propto s^{-q}$, where s is the linear size of the TNO and q is 4.0 ± 0.5 (refs 2, 3). Extrapolating this power-law distribution down to 20 m, which is roughly the smallest corresponding size of these dip events (see below), we obtain an occultation rate of about $4 \times 10^{-5} \text{ s}^{-1}$ for a background point

source. In the 322-ks exposure, about 12 occultation events can be expected. Our detection of 58 events is clearly more than that number but not by an order of magnitude.

The Fresnel scale, $R_F = (\lambda d / 2\pi)^{1/2}$, for a distance of $d = 40 \text{ AU}$ and an X-ray photon of $\lambda = 0.6 \text{ nm}$ (2 keV), is about 20 m, which is a significant fraction of the TNO size that we are considering. It is possible to resolve the degeneracy of TNO size and distance with the X-ray diffraction pattern in the occultation light curves, which are related to the size, shape and distance of the TNO as well as that of the Sco X-1 emission region. However, the various shapes of the occultation light curves need to be corrected for the dead-time and photon-coincidence effects, if they are to be studied further, because Sco X-1 is very bright for PCA¹³.

At present, we assume all the occulting TNOs are at a distance of 43 AU to estimate their size distribution. To convert event duration into TNO size, the relative projected speed on the sky of the RXTE with respect to the TNO in question at the corresponding epoch is needed. All the relative speeds and inferred TNO sizes of the 58 events are listed in Supplementary Table 2. Most of the events are due to TNOs about 50 m in size. As shown in Fig. 3, the inferred differential size distribution of TNOs of this size (Supplementary Discussion) is somewhat higher than the extrapolation from the best-fit distribution of TNOs larger than 100 km within 30–50 AU. We note that the discrepancy may still be within the uncertainty of that extrapolation, but numerical simulations of coagulation models for planet formation with collisional cascade predict a flatter distribution for smaller TNOs, with a turnover size between 0.1 and 100 km, depending on various properties of the early Solar System^{23,24}. Our results seem to favour a smaller turnover size. However, it is possible that some of the 58 TNOs are at distances beyond 50 AU. It is also possible that another component of smaller TNOs may exist, particularly if the reported Hubble Space Telescope estimate²⁵ of the size distribution at about 30 km is reliable; this estimate is a factor of 25 lower than expected from the above extrapolation and is based on the Hubble Space Telescope detection of three TNOs within 0.02 square degrees of the sky.

More RXTE observations of Sco X-1 are needed to improve the statistics of our current results. Future X-ray observations of Sco X-1 and other sources with new instruments of higher sensitivity could build on this discovery to shed further light on the formation of our Solar System.

Received 14 February; accepted 25 May 2006.

- Jewitt, D. C. & Luu, J. X. Discovery of the candidate Kuiper belt object 1992 QB₁. *Nature* **362**, 730–732 (1993).
- Trujillo, C. A., Jewitt, D. C. & Luu, J. X. Properties of the trans-Neptunian belt: Statistics from the Canada-France-Hawaii telescope survey. *Astron. J.* **122**, 457–473 (2001).
- Luu, J. X. & Jewitt, D. C. Kuiper belt objects: relics from the accretion disk of the sun. *Annu. Rev. Astron. Astrophys.* **40**, 63–101 (2002).
- Brown, M. E., Trujillo, C. A. & Rabinowitz, D. 2003 EL₆₁, 2003 UB₃₁₃, and 2005 FY₉. *IAU Circ.* **8577** (2005).
- Bertoldi, F., Altenhoff, W., Weiss, A., Menten, K. M. & Thum, C. The trans-neptunian object UB313 is larger than Pluto. *Nature* **439**, 563–564 (2006).
- Bailey, M. E. Can 'invisible' bodies be observed in the Solar System? *Nature* **259**, 290–291 (1976).
- Brown, M. J. I. & Webster, R. L. Occultations by Kuiper belt objects. *Mon. Not. R. Astron. Soc.* **289**, 783–786 (1997).
- Cooray, A. & Farmer, A. J. Occultation searches for Kuiper belt objects. *Astrophys. J.* **587**, L125–L128 (2003).
- Alcock, C. et al. TAOS: The Taiwanese-American Occultation Survey. *Earth Moon Planet* **92**, 459–464 (2003).
- Roques, F. et al. A search for small Kuiper belt objects by stellar occultations. *Astrophys. J.* **594**, L63–L66 (2003).
- Giacconi, R., Gursky, H., Paolini, F. & Rossi, B. Evidence for X-rays from sources outside the solar system. *Phys. Rev. Lett.* **9**, 439–443 (1962).
- Bradt, H. V., Rothschild, R. E. & Swank, J. H. X-ray timing explorer mission. *Astron. Astrophys. Suppl. Ser.* **97**, 355–360 (1993).
- Jahoda, K. et al. Peak luminosities of bursts from GRO J1744–28 measured with the RXTE PCA. *Nucl. Phys. B* **69**, 210–215 (1998).
- Middleditch, J. & Priedhorsky, W. C. Discovery of rapid quasi-periodic oscillations in Scorpius X-1. *Astrophys. J.* **306**, 230–237 (1986).

15. Scargle, J. D. *et al.* The quasi-periodic oscillations and very low frequency noise of Scorpius X-1 as transient chaos: a dripping handrail? *Astrophys. J.* **411**, L91–L94 (1993).
16. van der Klis, M., Wijnands, R., Horne, K. & Chen, W. KiloHertz quasi-periodic oscillation peak separation is not constant in Scorpius X-1. *Astrophys. J.* **481**, L97–L100 (1997).
17. Gierlinski, M. & Zdziarski, A. A. Discovery of powerful millisecond X-ray flares from Cygnus X-1. *Mon. Not. R. Astron. Soc.* **343**, L84–L88 (2003).
18. Uttley, P., McHardy, I. M. & Vaughan, S. Non-linear X-ray variability in X-ray binaries and active galaxies. *Mon. Not. R. Astron. Soc.* **359**, 345–362 (2005).
19. Steeghs, D. & Casares, J. The mass donor of Scorpius X-1 revealed. *Astrophys. J.* **568**, 273–278 (2002).
20. Bradshaw, C. F., Geldzahler, B. J. & Fomalont, E. B. The X-ray spectral changes of Scorpius X-1. *Astrophys. J.* **592**, 486–497 (2003).
21. Barnard, R., Church, M. J. & Balucinska-Church, M. Physical changes during Z-track movement in Sco X-1. *Astron. Astrophys.* **405**, 237–247 (2003).
22. Bradshaw, C. F., Fomalont, E. B. & Geldzahler, B. J. High resolution parallax measurements of Scorpius X-1. *Astrophys. J.* **512**, L121–L124 (1999).
23. Kenyon, S. J. Planet formation in the outer solar system. *Publ. Astron. Soc. Pacif.* **114**, 265–283 (2002).
24. Kenyon, S. J. & Bromley, B. C. The size distribution of Kuiper belt objects. *Astrophys. J.* **128**, 1916–1926 (2004).
25. Bernstein, G. L. *et al.* The size distribution of trans-Neptunian bodies. *Astrophys. J.* **128**, 1364–1390 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Cooray for comments that significantly improved this paper. We also benefited from discussions with R. Taam, Y.-H. Chu, W. Becker and C. Alcock. This research has made use of data obtained through the High Energy Astrophysics Science Archive Research Center Online Service, provided by the NASA/Goddard Space Flight Center, and of the JPL HORIZON online Solar System data and ephemeris computation service. This work was supported by the National Science Council of the Republic of China and Academia Sinica in Taipei.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.-K.C. (hkchang@phys.nthu.edu.tw).

LETTERS

Tracking the motion of charges in a terahertz light field by femtosecond X-ray diffraction

A. Cavalleri^{1,2}, S. Wall¹, C. Simpson¹, E. Statz³, D. W. Ward³, K. A. Nelson³, M. Rini⁴ & R. W. Schoenlein⁴

In condensed matter, light propagation near resonances is described in terms of polaritons, electro-mechanical excitations in which the time-dependent electric field is coupled to the oscillation of charged masses^{1,2}. This description underpins our understanding of the macroscopic optical properties of solids, liquids and plasmas, as well as of their dispersion with frequency. In ferroelectric materials, terahertz radiation propagates by driving infrared-active lattice vibrations, resulting in phonon-polariton waves. Electro-optic sampling with femtosecond optical pulses^{3–5} can measure the time-dependent electrical polarization, providing a phase-sensitive analogue to optical Raman scattering^{6,7}. Here we use femtosecond time-resolved X-ray diffraction^{8–10}, a phase-sensitive analogue to inelastic X-ray scattering^{11–13}, to measure the corresponding displacements of ions in ferroelectric lithium tantalate, LiTaO₃. Amplitude and phase of all degrees of freedom in a light field are thus directly measured in the time domain. Notably, extension of other X-ray techniques to the femtosecond timescale (for example, magnetic or anomalous scattering) would allow for studies in complex systems, where electric fields couple to multiple degrees of freedom¹⁴.

Below $T_c \approx 900^\circ\text{C}$, LiTaO₃ is ferroelectric, with a permanent *c*-axis structural distortion coupled to a permanent electric dipole. At terahertz frequencies, light couples with periodic lattice distortions around the ferroelectric equilibrium positions, resulting in a phonon-polariton mode of A₁ symmetry. Because of the reduced symmetry of the ferroelectric phase, LiTaO₃ has also a large optical nonlinearity, and consequently femtosecond visible pulses can be used to generate terahertz radiation by means of impulsive stimulated Raman scattering (ISRS). The resulting time-dependent electrical polarization of coherent phonon polaritons has been measured in a number of experiments with electro-optic sampling, spatiotemporal imaging and other optical observables at visible frequencies. However, the corresponding lattice motions have to date remained undetected, leading sometimes to controversy on the interpretation of coherent polariton experiments¹⁵. Here we use femtosecond X-rays to measure the corresponding atomic oscillations¹⁶, which we compare with a time-domain simulation of terahertz propagation to extract the absolute displacements with 1-mÅ resolution.

Figure 1a shows a sketch of the experimental geometry. Coherent phonon polaritons in LiTaO₃ were excited nonlinearly at the surface by a 70-fs, *s*-polarized, 800-nm laser pulse at a fluence of 10 mJ cm^{-2} . Terahertz lattice distortions were then measured directly as time-dependent modulations of the 006 structure factor. To maximize the temporal resolution of our experiments, the 68° incidence angle for both pump and probe (22° according to the X-ray convention) was then dictated by the Bragg condition for the 7-keV X-ray probe. A finite-difference time-domain simulation, describing the excitation and propagation of phonon-polaritons¹⁷ at arbitrary angles, was used

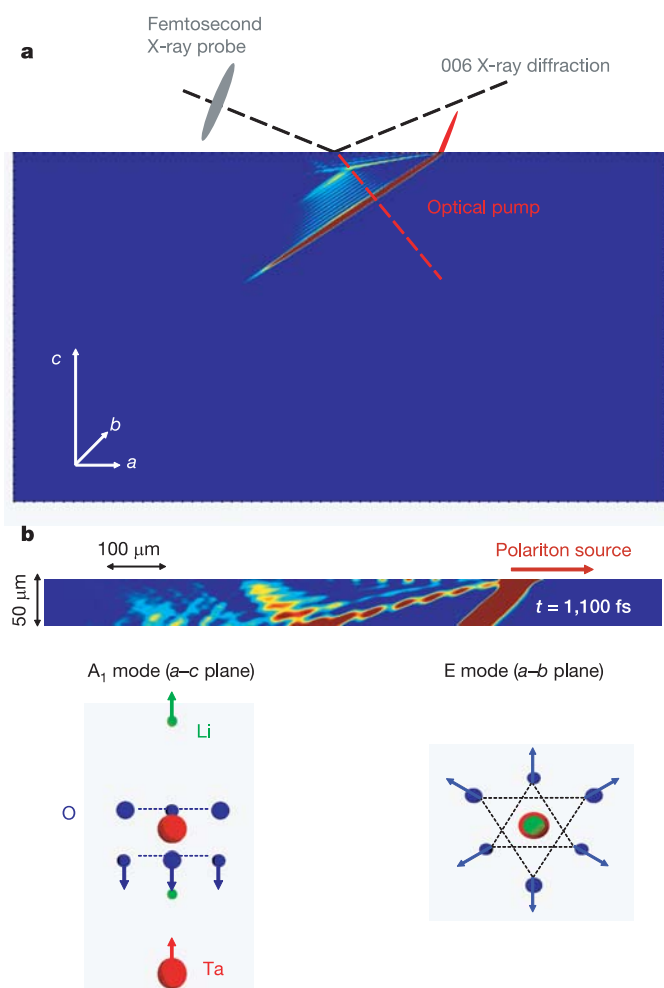


Figure 1 | Excitation of phonon polaritons in LiTaO₃. **a**, Experimental geometry. Phonon-polaritons were excited via ISRS with *s*-polarized, 70-fs pulses of near-800-nm wavelength at a fluence of 10 mJ cm^{-2} . The optical pump pulse impinged onto the sample at the 68° angle, as dictated by the Bragg condition for the 7-keV, X-ray diffraction probe. The pump pulse was thus refracted at a 32° angle and generated a time- and space-dependent pattern of terahertz phonon-polaritons. The lattice distortions along the *c* axis of LiTaO₃ were measured as modulations of the 006 structure factor. **b**, Snapshot of the spatial pattern of polaritons. The figure shows a $40\text{-}\mu\text{m}$ -wavelength surface wave with propagation along the surface normal. The measured lattice distortions associated with the A₁ and E modes are shown. The polarization component in the *a*–*b* plane, associated with a vibrational mode of E symmetry, is not detected in this reflection.

¹Department of Physics, Clarendon Laboratory, University of Oxford, Oxford OX1 3PU, UK. ²Central Laser Facility & Diamond Light Source, Rutherford Appleton Laboratory, Chilton, Didcot, OX11 0QX, UK. ³Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁴Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

to predict and interpret the experimental results. In our simulations, the terahertz radiation field was coupled linearly to the lattice vibrations, which were treated as a driven ionic oscillator. The relevant lattice vibrations and their coupling to electromagnetic radiation were described through an auxiliary parameter, which accounted for the potential-energy surface of the specific phonon-mode. Lattice vibrations contributed to the terahertz radiation field through the ionic polarization, which was linear in the ionic displacement. Polariton generation with femtosecond optical pulses of arbitrary time duration and fluence was described by an additional driving term to the ionic oscillator that is proportional to the ISRS excitation intensity.

A snapshot of the simulated phonon-polariton pattern, calculated at first with a coarse 5- μm spatial mesh, is also displayed in Fig. 1a. The 800-nm pump pulse was refracted at 32° from the surface normal and two polariton waves propagated at a slower group velocity in the wake of the laser pulse. The first terahertz wave followed at the same 32° angle of the optical pump, while a second one propagated at 8° from the surface normal.

A higher-resolution simulation, reported in Fig. 1b, explains the excitation mechanism. Refraction of the 800-nm pump pulse over the 1-mm² spot resulted in an optical wave front that crossed the vacuum-material interface at different points for different time delays. The 21- μm length of the 70-fs optical pulse corresponded then to a nonlinear 'source' that travelled along the surface at an effective speed of $1.08c_0$, where c_0 is the speed of light in vacuum. This speed is only apparent, and does not imply superluminal propagation.

In the spirit of the Huygen's principle, and keeping into account both linear (800-nm) and nonlinear (terahertz) terms, the re-radiated field can be broken into three components: (1) The linearly re-radiated optical field propagates into the solid at $c_0/2.25$, where $n \approx 2.25$ is the index of refraction of LiTaO₃ at 800 nm. This effect results in conventional refraction at 32° from the surface normal. (2) The terahertz waves generated beneath the surface followed the optical pump-pulse at the same 32° angle. (3) A second wave at terahertz frequencies was also generated at the very surface, resulting from a moving nonlinear source that was matched to a wave propagating into the solid at a speed of $c_0/6.1$ ($n \approx 6.1$), resulting in propagation at 8° from the surface normal. The 40- μm wavelength for the terahertz radiation was then dictated by the projection onto the a - b plane of the longitudinal width of the optical pulse ($\sim 21 \mu\text{m}/\cos 68^\circ$). Thus, a 1.5-THz wave entered the solid as if it were

impinging from vacuum collinearly with the optical pulse and as if it were refracted linearly into the material.

Two polarization components were excited in the terahertz field, one parallel and one perpendicular to the optic axis of LiTaO₃. These were associated with two different vibrational modes of the crystal, one of A_1 symmetry along the c (or optic) axis, and one of E symmetry in the a - b plane. Figure 1b shows the atomic displacements for these two normal modes. Diffraction from the 006 plane was insensitive to all distortions of the unit cell in the a - b plane, which can only be detected for nonzero h and/or k (two of the three Miller indices, for an hkl X-ray reflection). Thus, our experiment could only detect distortions of A_1 symmetry along the c axis.

Optical pump, X-ray probe experiments were performed by using synchrotron radiation from laser-modulated electron bunches^{18–20}, which interacted with the optical field within a wiggler of the storage ring (see Fig. 2). Two energy sidebands at approximately 10 MeV from the 1.9-GeV central energy of the beam were impressed by interaction of the electron beam with the optical field. Femtosecond electron bunches were then transiently separated from the main orbit in the next bending magnet, where the electron beam energy was dispersed along the radius of the storage ring. Femtosecond synchrotron radiation was thus imaged onto the sample using a toroidally bent silicon mirror, and resulted in X-ray pulses that were separated by approximately 500 μm from the main axis of the beam. Broad-band, femtosecond X-rays were passed through a two-crystal Si-111 monochromator, producing a train of monochromatic X-ray pulses at 7 keV. The spot size of the probing X-rays was of approximately $200 \mu\text{m} \times 150 \mu\text{m}$, projected over a surface region approximately 300- μm long, due to the 68° angle of incidence (measured from the surface normal). The detected flux in the femtosecond X-rays was approximately four orders of magnitude lower than in the picosecond pulse, amounting to a few thousand photons per second of 0.1% bandwidth in a 2-kHz train. After diffraction, a few hundred photons per second were detected by the avalanche photo diode (APD). A cross-correlation between the pump laser and the sliced

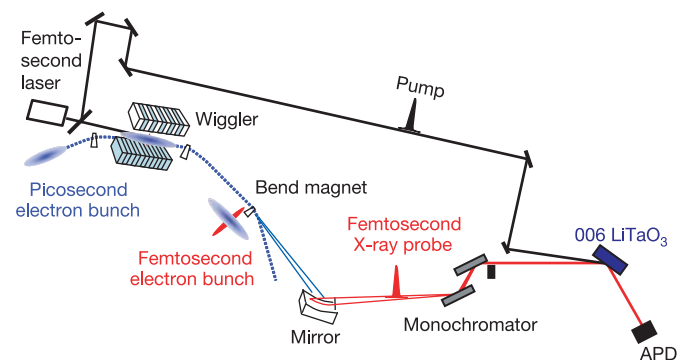


Figure 2 | Optical pump, X-ray diffraction probe experimental apparatus. A single Ti:Sa laser oscillator was synchronized to the storage-ring orbit clock to about 1 ps. These low-energy pulses were amplified in two separate regenerative amplifiers. The laser pulses used to impress energy modulation onto the electron bunches and those used to excite the sample were derived from the same oscillator, ensuring absolute synchronization between optical pump and X-ray probe. Time-dependent modulations of the integrated 006 diffraction peak were measured at 7 keV as a function of the pump-probe time delay.

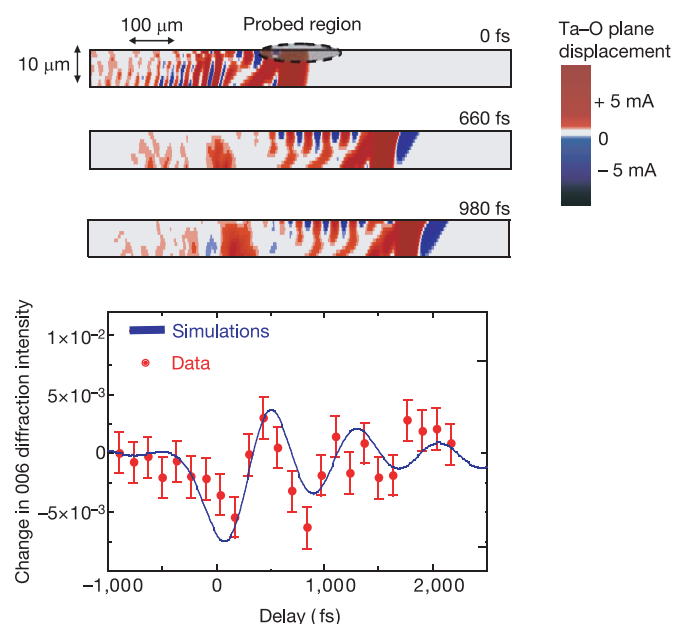


Figure 3 | Measurement of the time-dependent change in the 006 diffracted intensity. The error bars in the time-dependent X-ray diffraction data represent one standard deviation of the shot-noise-limited distribution of detected photons. In the three simulated patterns, the colour code represents the absolute displacement of the Ta atom with respect to the plane of oxygen atoms, taking positive and negative values with respect to the equilibrium (distorted) positions. The continuous curve in the lower panel is then a one-parameter fit to the measured experimental data.

pulses of visible bending magnet radiation revealed a pulse duration of approximately 200 fs.

Figure 3 displays the measured time-dependent scattering intensity. A 1.5-THz modulation of the 006 structure factor was found, with a sine-like time dependence that reflects the electronic-ground-state nature of lattice motions initiated with ISRS. The data were compared to the results of the polariton propagation simulation, which were performed with submicrometre resolution and are also shown in Fig. 3. The colour code of the polariton pattern represents the amplitude of the *c*-axis displacements of the A_1 mode, highlighting positive and negative motion of the Ta with respect to the oxygen plane. The sub-surface polariton patterns were combined with the time- and space-dependent displacements of the A_1 normal mode of Fig. 1b, and were integrated spatially over the probed volume to account for the change in the 006 X-ray reflectivity. The continuous curve superimposed on the data displays the results of our simulation. Importantly, because the extinction depth was of less than 1 μm , our femtosecond X-ray diffraction experiments sampled only the polariton wave immediately beneath the surface. Because the probed near-surface volume was homogeneously excited at all times, dynamic diffraction simulations gave identical results to those obtained in the kinematic approximation.

No free parameter was left in the calculation, except for the absolute displacements of the A_1 normal mode, which were adjusted to fit the data. The colour code of Fig. 3 shows a 5-mÅ peak motion of the Ta atom with respect to the plane of the oxygen atoms. This motion corresponds to approximately 3% of what is needed to reach the ferroelectric phase at higher temperatures ($T_c = 900\text{ K}$).

In summary, we used femtosecond X-ray diffraction to measure the lattice motions that sustain terahertz light propagation in ferroelectric LiTaO_3 . Ultrafast lattice motions along the A_1 lattice direction were driven by phonon-polaritons excited at the solid–vacuum interface by ISRS. Our experiments and simulations reveal a polariton wave with a frequency of 1.5 THz, as set by the pulse duration and angle of incidence of our optical pulse. Similar experiments with improved X-ray focusing capabilities would allow for the imaging of space- and time-dependence in this and similar problems, allowing for full mapping of coherently excited lattice waves²¹. Femtosecond X-ray measurements could also be coupled to high-amplitude lattice excitation, to drive the structure into transient or permanent states that are not accessed near equilibrium.

More generally, ultrafast X-ray diffraction from terahertz lattice waves opens the way to new dynamic studies in complex solids in their electronic ground state, in which microscopic couplings among multiple degrees of freedom dominate the physics^{22–24}. Finally, the possibility of controlling propagating phonon-polariton modes may result in new concepts for ultrafast Bragg switches^{25,26}, capable of separating femtosecond slices from longer X-ray pulses.

Received 1 May; accepted 30 June 2006.

1. Mills, D. L. & Burstein, E. Polaritons: the electromagnetic modes of media. *Rep. Prog. Phys.* **37**, 817–926 (1974).
2. Barker, A. S. Jr & Loudon, R. Response functions in the theory of Raman scattering by vibrational and polariton modes in dielectric crystals. *Rev. Mod. Phys.* **44**, 18–47 (1972).
3. Auston, D. H., Cheung, K. P., Valdamanis, J. A. & Kleinman, D. Cherenkov radiation from femtosecond optical pulses in electro-optic media. *Phys. Rev. Lett.* **53**, 1555–1558 (1984).
4. Wahlstrand, J. K. & Merlin, R. Cherenkov radiation emitted by ultrafast laser pulses and the generation of coherent polaritons. *Phys. Rev. B* **68**, 054301 (2003).

5. Crimmins, T. F., Stoyanov, N. S. & Nelson, K. A. Heterodyned impulsive stimulated Raman scattering of phonon-polaritons in LiTaO_3 and LiNbO_3 . *J. Chem. Phys.* **117**, 2882–2896 (2002).
6. Dougherty, T. P. et al. Femtosecond resolution of soft mode dynamics in structural phase transitions. *Science* **258**, 770–774 (1992).
7. Stevens, T. E., Wahlstrand, J. K., Kuhl, J. & Merlin, R. Cherenkov radiation at speeds below the light threshold: phonon assisted phase matching. *Science* **291**, 627–630 (2001).
8. Rischel, C. et al. Femtosecond time-resolved X-ray diffraction from laser heated organic films. *Nature* **390**, 490–492 (1997).
9. Cavalleri, A. et al. Femtosecond structural dynamics in VO_2 during an ultrafast solid-solid phase transition. *Phys. Rev. Lett.* **87**, 237401 (2001).
10. Lindenberg, A. M. et al. Atomic scale visualization of inertial dynamics. *Science* **308**, 392–395 (2005).
11. Cavalleri, A. et al. Anharmonic lattice dynamics in germanium measured with ultrafast x-ray diffraction. *Phys. Rev. Lett.* **85**, 586–589 (2000).
12. Lindenberg, A. M. et al. Time-resolved X-ray diffraction from coherent phonons during a laser-induced phase transition. *Phys. Rev. Lett.* **84**, 111–114 (2000).
13. Bargheer, M. et al. Coherent atomic motions in a nanostructure studied by femtosecond x-ray diffraction. *Science* **306**, 1771–1773 (2004).
14. Kimel, A. V. et al. Ultrafast non-thermal control of magnetization by instantaneous photo-magnetic pulses. *Nature* **435**, 655–657 (2005).
15. Bakker, H. J., Hunsche, S. & Kurz, H. Time-resolved study of phonon polaritons in LiTaO_3 . *Phys. Rev. B* **48**, 13524–13537 (1993).
16. Sokolowski-Tinten, K. et al. Femtosecond X-ray measurement of coherent lattice vibrations near the Lindemann stability limit. *Nature* **422**, 287–289 (2003).
17. Stoyanov, N. S., Ward, D. W., Feurer, Th. & Nelson, K. A. Terahertz polariton propagation in patterned materials. *Nature Mater.* **1**, 95–98 (2002).
18. Zholents, A. A. & Zolotarev, M. S. Femtosecond X-ray pulses of synchrotron radiation. *Phys. Rev. Lett.* **76**, 912–915 (1996).
19. Schoenlein, R. W. et al. Generation of femtosecond pulses of synchrotron radiation. *Science* **287**, 2237–2240 (2000).
20. Cavalleri, A. et al. Band selective measurements of electron dynamics in VO_2 using femtosecond near edge X-ray absorption. *Phys. Rev. Lett.* **95**, 067405 (2005).
21. Feurer, T., Vaughan, J. C. & Nelson, K. A. Spatiotemporal coherent control of lattice vibrational waves. *Science* **299**, 374–377 (2003).
22. Kimura, T. et al. Magnetic control of ferroelectric polarization. *Nature* **426**, 55–58 (2003).
23. Lottermoser, Th. et al. Magnetic phase control by an electric field. *Nature* **430**, 541–544 (2004).
24. Kimel, A. V., Kirilyuk, A., Tsvetkov, A., Pisarev, R. V. & Rasing, Th. Laser-induced ultrafast spin re-orientation in the antiferromagnet TmFeO_3 . *Nature* **429**, 850–853 (2004).
25. Bucksbaum, P. H. & Merlin, R. The phonon Bragg switch: a proposal to generate sub-picosecond x-ray pulses. *Solid State Commun.* **111**, 535–539 (1999).
26. Nazarkin, A., Uschman, I., Förster, E. & Sauerbrey, R. High order Raman scattering of X-rays by optical phonons and generation of ultrafast X-ray transients. *Phys. Rev. Lett.* **93**, 207401 (2004).

Acknowledgements We thank M. Khalil for her help during data acquisition. We are grateful to J. S. Wark for discussions, as well as for sharing the results of dynamic diffraction simulations for LiTaO_3 . We thank K. Sokolowski-Tinten, R. Merlin, D. Reis and S. Hooker for many discussions, suggestions and for critical reading of our manuscript. Help by N. Tamura with the measurement of static Laue patterns in LiTaO_3 is also gratefully acknowledged. Experiments at LBNL were supported by the Director, Office of Science, Office of Basic Energy Sciences, Division of Materials Science and Engineering Division, of the US Department of Energy. Work at the University of Oxford was supported by the European Science Foundation through a European Young Investigator Award. S.W. acknowledges receipt of a graduate scholarship from the UK Engineering and Physical Sciences Research Council. Simulation work at MIT was funded by the US National Science Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.C. (a.cavalleri1@physics.ox.ac.uk).

Supercurrent reversal in quantum dots

Jorden A. van Dam¹, Yuli V. Nazarov¹, Erik P. A. M. Bakkers², Silvano De Franceschi^{1,3} & Leo P. Kouwenhoven¹

When two superconductors are electrically connected by a weak link—such as a tunnel barrier—a zero-resistance supercurrent can flow^{1,2}. This supercurrent is carried by Cooper pairs of electrons with a combined charge of twice the elementary charge, e . The $2e$ charge quantum is clearly visible in the height of voltage steps in Josephson junctions under microwave irradiation, and in the magnetic flux periodicity of $h/2e$ (where h is Planck's constant) in superconducting quantum interference devices³. Here we study supercurrents through a quantum dot created in a semiconductor nanowire by local electrostatic gating. Owing to strong Coulomb interaction, electrons only tunnel one-by-one through the discrete energy levels of the quantum dot. This nevertheless can yield a supercurrent when subsequent tunnel events are coherent^{4–7}. These quantum coherent tunnelling processes can result in either a positive or a negative supercurrent, that is, in a normal or a π -junction^{8–10}, respectively. We demonstrate that the supercurrent reverses sign by adding a single electron spin to the quantum dot. When excited states of the quantum dot are involved in transport, the supercurrent sign also depends on the character of the orbital wavefunctions.

The electronic properties of quantum dots can be probed by attaching source and drain electrodes, allowing charge carriers to tunnel from the dot to both electrodes. If the electrodes are superconducting, transport is strongly affected and largely depends on the transparency of the electrical connection between the electrodes and the quantum dot. A number of experiments have focused on various phenomena in the Coulomb blockade regime but no supercurrents through quantum dots were observed, mostly due to the lack of a controllable tunnel coupling with the electrodes^{11–14}. Strong coupling and negligible Coulomb interactions were recently obtained in carbon nanotube quantum dots demonstrating resonant tunnelling of Cooper pairs through a single quantum state¹⁵. In the regime of strong Coulomb interactions, the simultaneous occupation of the quantum dot with two electrons is unfavourable. Nevertheless, a supercurrent can flow owing to the subsequent (but coherent) transport of correlated electrons. This can give rise to a sign change of the Cooper pair singlet (that is, from $(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}$ to $e^{i\pi}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}$, where $\uparrow$ and $\downarrow$ represent the two eigenstates of the z component of the spin). Therefore, the typical Josephson relation between the supercurrent, I_s , and the macroscopic phase difference between the superconductors, φ , usually given by $I_s = I_c \sin(\varphi)$, changes to $I_s = I_c \sin(\varphi + \pi) = -I_c \sin(\varphi)$ (ref. 5; I_c is the critical current). Other mechanisms of Cooper pair transport resulting in negative supercurrents have been studied using high-transition-temperature superconductors⁸, ferromagnets⁹, and non-equilibrium mesoscopic normal metals¹⁰.

We use indium arsenide (InAs) nanowires as semiconductor weak links¹⁶ in combination with local gate electrodes in order to obtain quantum dots with a tunable coupling to superconducting leads. The monocrystalline n-type InAs nanowires are grown by a catalytic process based on the vapour–liquid–solid growth method^{17–20}. After

growth, the wires are transferred to an oxidized silicon substrate. Previously developed nanofabrication techniques are used to define highly transparent aluminium-based superconducting contacts¹⁶. Pairs of nearby nanowires are contacted in parallel, forming a superconducting loop with two nanowire junctions (Fig. 1a). In a second lithographic step, we define local gate electrodes. One of the nanowires (top nanowire in Fig. 1a) is crossed by two gates, labelled L and R, in order to define a quantum dot (also see Fig. 1b). The bottom nanowire is crossed by one gate, labelled REF, and will be used as a reference junction with a tunable Josephson coupling. We have studied two similar devices in detail. Here we present the results for one of them. Similar data from the second device and further details on device fabrication are given as Supplementary Information.

Below the superconducting transition temperature of the aluminium-based contacts ($T_c \approx 1.1$ K), the two nanowires form superconducting weak links owing to the proximity effect¹⁶, thereby realizing a superconducting quantum interference device (SQUID)². The critical current of the SQUID, I_c , as a function of magnetic flux, Φ , shows oscillations with a period of $66 \mu\text{T}$. This is consistent with the addition of a flux quantum, $\Phi_0 = h/2e$, to the effective SQUID area of $30 \mu\text{m}^2$ (Fig. 1c, blue trace, temperature $T = 30$ mK). The maximum (minimum) critical current corresponds to the sum (difference) of the critical currents of the two nanowire junctions. Unlike in other SQUIDs, the critical currents of the individual junctions can be tuned by applying voltages to the respective gates. This is demonstrated by a measurement of the SQUID oscillations for different voltages applied to REF. When $V_{\text{REF}} = -0.64$ V (Fig. 1c, green trace), the amplitude of the SQUID oscillations is reduced owing to the partial local depletion of the nanowire. By further reducing the gate voltage to $V_{\text{REF}} = -0.80$ V, the reference junction is pinched off, resulting in the disappearance of the interference signal. We thus have a unique electrical control over the SQUID operation.

A quantum dot is formed in the top nanowire by applying negative voltages simultaneously to gates L and R. The local depletion creates two tunnel barriers, which define a single quantum dot in the nanowire section between the gates (see Fig. 2a inset), giving rise to discrete energy levels and Coulomb blockade. To show this, we pinch off the reference junction ($V_{\text{REF}} = -0.80$ V) and apply a small magnetic field in order to suppress superconductivity. Figure 1d shows a colour plot of absolute current through the quantum dot, $|I|$, as a function of bias voltage, V , and gate voltages, $V_L = V_R$. Coulomb blockade ($|I| = 0$) occurs within continuous diamond-shaped regions, as is typically observed in transport through single quantum dots²¹. Outside these regions, $|I|$ increases in steps (lines parallel to the diamond edges) denoting the onset of single-electron tunnelling via discrete excited states. From the separation between these lines, we estimate for this regime a characteristic level spacing of ~ 1 meV. The sharpness of the diamond edges and the excitation lines denote a weak tunnel coupling between the quantum dot and the source and drain leads. We can increase the coupling by reducing the negative

¹Kavli Institute of Nanoscience, Delft University of Technology, PO Box 5046, 2600 GA, Delft, The Netherlands. ²Philips Research Laboratories, High Tech Campus 4, 5656 AE Eindhoven, The Netherlands. ³Laboratorio Nazionale TASC INFN-CNR, Area Science Park, S.S. 14, Km. 163.5, I-34012 Trieste, Italy.

voltages applied to L and R (see Fig. 1e). This results in blurred diamond edges (dotted lines) and the appearance of inelastic co-tunnelling features inside the diamonds. This tunable coupling is particularly important for reaching the narrow transport regime where charging effects dominate but, at the same time, the critical current is large enough to be measurable.

Switching to the superconducting state but with the reference junction still pinched off, two peaks in dI/dV develop around $V \approx \pm 200 \mu\text{V} = \pm 2\Delta^*/e$ (Fig. 2b). $2\Delta^*$ is the superconducting gap induced in the nanowire by the proximity effect (Fig. 2a inset). These features are due to second-order co-tunnelling, and the peak shape reflects the singularities in the quasiparticle density of states at the gap edges. In spite of the Coulomb blockade effect, we observe a finite supercurrent, $I_{c,qd}$, through the nanowire quantum dot. We exploit the SQUID geometry to determine the critical value and the sign of this supercurrent in a current-biased measurement²². When an integer number of flux quanta are applied through the SQUID area, the critical current of the SQUID corresponds to the sum of the critical currents of the two junctions², that is, $I_c = I_{c,qd} + I_{c,REF}$. We set $I_{c,REF} = 320 \text{ pA}$, and extract the V_L -dependence of $I_{c,qd}$ directly

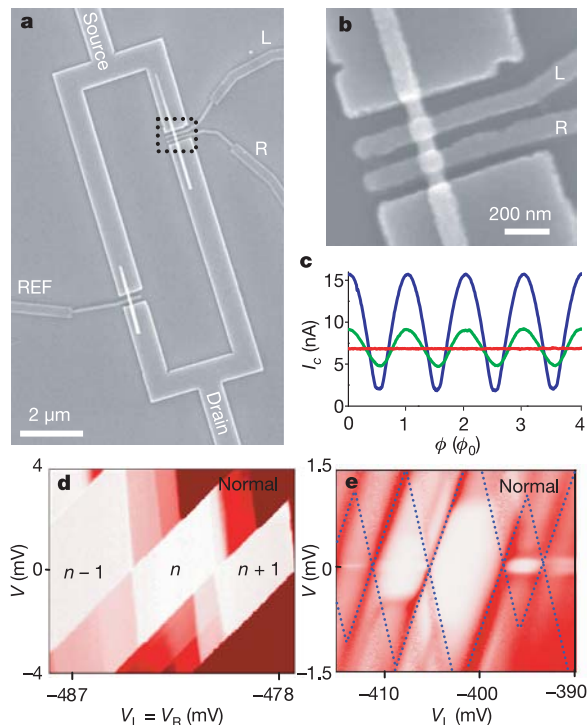


Figure 1 | Sample layout and device characterization. **a**, Scanning electron micrograph of the InAs nanowire SQUID. Two nanowires (diameter $\sim 60 \text{ nm}$) are incorporated in a superconducting loop (100 nm Al on 10 nm Ti). Aluminium top-gates (L and R) with a spacing of $\sim 65 \text{ nm}$ are used to define a quantum dot in the top nanowire. A third gate (REF) is used to control the reference junction. **b**, High-resolution image of the top nanowire shown boxed in **a**. **c**, Critical current of the SQUID, I_c , versus magnetic flux, Φ (in units of the flux quantum, Φ_0), for different voltages applied to the reference gate ($V_{REF} = 0 \text{ V}$ (blue trace), -0.64 V (green), and -0.80 V (red)), demonstrating full electrical control over the amplitude of the SQUID oscillations. **d**, Colour plot of absolute current through the dot, $|I|$, (increasing from white (0 pA) to red (5 pA)) versus source-drain bias voltage, V , and $V_L = V_R$ in the normal state. The Coulomb diamonds are well defined owing to the weak tunnel coupling between quantum dot and leads. **e**, Differential conductance, dI/dV (increasing from white (0.1 μS) to red (40 μS)), as a function of V and V_L ($V_R = -0.40 \text{ V}$). The stronger dot-lead coupling results in blurred diamond edges (indicated by dotted lines) and horizontal features inside the diamonds due to inelastic co-tunnelling. Data in **d** and **e** are taken at $T = 30 \text{ mK}$, and in a small magnetic field to drive the superconducting contacts into the normal state.

from the measurement of I_c (Fig. 2a). We find $I_{c,qd} < 0$ for two charge states of the quantum dot, denoted by a yellow diamond and a yellow square in Fig. 2b. The negative supercurrent of the quantum dot junction is confirmed by the $\Phi_0/2$ shift between the SQUID oscillations for $I_{c,qd} < 0$ (Fig. 2c, red trace) and those for $I_{c,qd} > 0$ (blue trace). A colour plot of $I_c(V_L, \Phi)$ in Fig. 2d shows the transitions between positive and negative supercurrents around the charge state denoted by the yellow square.

Negative supercurrents have been predicted for superconductors coupled by a magnetic impurity or a single-level interacting quantum dot^{3–5,7}. In these systems resonant tunnelling of Cooper pairs is prohibited owing to Coulomb blockade. Nevertheless, Cooper pairs can be transported via fourth-order co-tunnelling events. Three examples of such events are shown in Fig. 3. The top and bottom diagrams are the initial and final states, respectively, and the

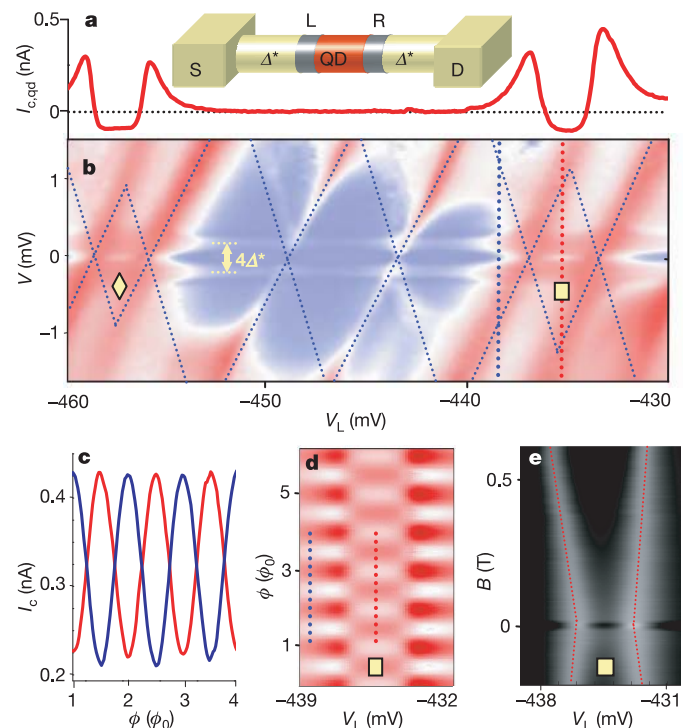


Figure 2 | Supercurrent reversal in an interacting quantum dot. **a**, Plot of the critical current of the quantum dot, $I_{c,qd}$, as a function of gate voltage, V_L , for the same gate voltage region as in **b**. QD, quantum dot; S, source; D, drain. A negative supercurrent is observed for two charge states. Inset, diagram of the quantum dot in the nanowire. **b**, Colour scale plot of differential conductance, $dI/dV(V, V_L)$, in the superconducting state (dI/dV increases from blue, through white, to red; $V_{REF} = -0.8 \text{ V}$). The two peaks in dI/dV at $V \approx \pm 200 \mu\text{V}$ are due to quasiparticle co-tunnelling, and their spacing ($4\Delta^*$ as indicated) provides a direct measurement of the induced superconducting gap in the nanowire. Yellow symbols (diamond and square) indicate two charge states that exhibit negative supercurrent. Blue dotted lines indicate the diamond edges. **c**, Two $I_c(\Phi)$ curves taken at gate voltages indicated by the vertical red and blue dotted lines in **b**, demonstrating the shift by $\Phi_0/2$ between the conventional (blue) and the π -regime (red). **d**, Critical current of the SQUID, I_c , in colour-scale as a function of Φ and V_L . Yellow square: the interference signal is shifted by half a flux quantum compared to adjacent Coulomb diamonds, indicating the π -shift in the Josephson relation. Red and blue dotted lines correspond to red and blue traces in **c**. Individual $I_c(\Phi)$ curves can be fitted very well with a sine-function within the measurement accuracy. **e**, Grey-scale plot of linear conductance, G (increasing from black to white), as a function of magnetic field, B , and V_L . Yellow square: the Coulomb peak spacing in this charge state increases with increasing field owing to the Zeeman effect, indicating that the number of electrons is odd. All measurements are taken at $T = 30 \text{ mK}$. Note that the measurements in **a**, **c** and **d** are current-biased and in **b** and **e** voltage-biased.

diagrams in between show one of the three intermediate virtual states. Owing to Coulomb blockade, a sequence of intermediate states involves an energy cost comparable to the charging energy, E_c (for $\Delta^* \ll E_c$). Nevertheless, when the tunnel rate is on the order of $E_c/\hbar$, a Cooper pair can be transported by higher-order co-tunnelling events²³. In principle, there are 24 possible sequences of 4 tunnel events. However, in a single-level quantum dot only a small number of sequences are allowed. Figure 3a illustrates the transfer of a Cooper pair through a quantum dot with a single spin-degenerate level occupied by one electron (with spin up, $|\uparrow\rangle$). The sequence of four tunnel processes, indicated by the numbers, is necessarily permuted compared to ordinary transport of Cooper pairs. The remarkable result is that the spin-ordering of the Cooper pair is reversed, that is, the Cooper pair on the right is created in the order $|\uparrow\rangle, |\downarrow\rangle$ while the pair on the left is annihilated in the order $|\downarrow\rangle, |\uparrow\rangle$. This spin-reversal results in a sign-change of the Cooper pair singlet state (for example, from $(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}$ to $e^{i\pi}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}$), leading to a π -shift in the Josephson relation and a negative supercurrent. However, if an extra electron is added to the quantum dot, the sequence of tunnel events discussed above is prohibited, owing to the Pauli exclusion principle. Now other sequences of tunnel events are allowed, which result in a normal, positive supercurrent⁷ (Fig. 3b). Therefore, in a single-level quantum dot a negative (positive) supercurrent is expected for an odd (even) number of electrons.

We can discriminate between odd and even numbers of electrons in Fig. 2b by measuring the linear conductance, G , as a function of gate voltage and magnetic field, B (Fig. 2e). We observe that the Coulomb peak spacing for the two charge states denoted by the yellow square and diamond increases owing to the Zeeman effect,

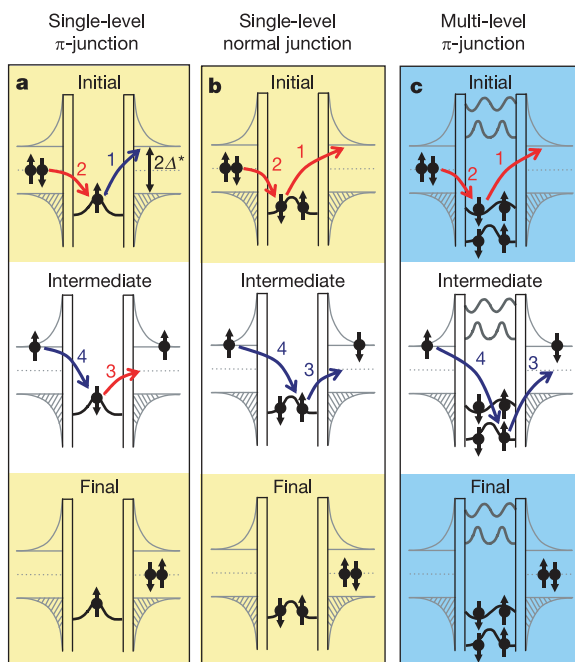


Figure 3 | Energy diagrams illustrating Cooper pair transport through a quantum dot due to fourth-order co-tunnelling. Top and bottom panels represent initial and final states, respectively. The intermediate panels show one of the three virtual intermediate states. Numbers indicate the sequence of tunnel events. Red (blue) corresponds to the tunnelling of a spin-down (spin-up) electron. **a**, Transport occurs through a single spin-degenerate level filled with one electron. During this event the spin-ordering of the Cooper pair is reversed. This results in a negative contribution to the supercurrent (see also diagrams in ref. 5). **b**, Transport through one spin-degenerate level filled with two electrons. The spin-ordering of the Cooper pair cannot be reversed, resulting always in a positive supercurrent. **c**, Co-tunnelling event involving two energy levels with wavefunctions of opposite parity. This results in a negative contribution to the supercurrent⁷.

demonstrating that for these charge states the occupation number, n , is odd²¹ (only data for the state denoted by a yellow square is shown, $|g\text{-factor}| \approx 15$, similar to previous results for similar systems²⁴). These observations are consistent with the model described above.

However, for the charge state around $V_L = -447$ mV with an odd number of electrons, we observe a very small, but positive critical current ($I_c \approx 10$ pA). Moreover, in a different gate voltage range, shown in Fig. 4a, supercurrent reversal is also observed for charge states with an even number of electrons. We argue that these observations originate from co-tunnelling via multiple energy levels of the quantum dot. The multi-level nature of the quantum dot for the gate range studied in Fig. 4a emerges from the measurement of differential conductance in the normal state (Fig. 1e). Here several peaks parallel to the diamond edges are observed, which correspond to transport through excited states of the quantum dot. In this gate voltage range the level spacing, δ , is of the order of E_c . Therefore, these excited states can take part in co-tunnelling events and the simple model of a single-level quantum dot is no longer appropriate. As a result, all 24 sequences of tunnel events are allowed for both odd and even numbers of electrons. Therefore, a negative supercurrent due to permutation of tunnel events is possible for all values of n (refs 7, 25).

Additionally, in the multi-level regime, properties of the wavefunctions of the quantum dot become important. To illustrate this, we consider the co-tunnelling event in Fig. 3c in which two different energy levels are involved in a dot with an even number of electrons. Because the two electrons take a different path, they can acquire a different phase. The opposite parity of the wavefunctions results in a

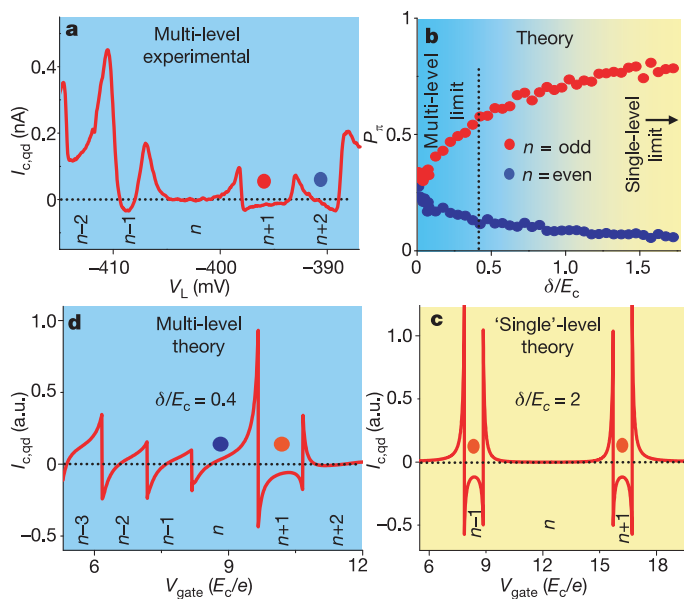


Figure 4 | Experimental results and numerical simulations for a multi-level quantum dot. Panels are ordered clockwise. **a**, Measured critical current of the quantum dot, $I_{c,qd}$, as a function of V_L , showing supercurrent reversal for even and odd numbers of electrons (indicated by a blue and red dot, respectively; $V_R = -0.4$ V). **b**, Calculated probability of π -behaviour, P_π , for odd (red) and even (blue) numbers of electrons as a function of δ/E_c . Strength and sign of tunnel couplings are randomly varied. For $\delta/E_c \gg 1$, the limiting case of a single-level quantum dot is reached, resulting in π -behaviour for odd numbers and conventional behaviour for even numbers of electrons. In the multi-level limit ($\delta/E_c \ll 1$), we obtain $P_\pi \approx 0.3$ for both even and odd numbers of electrons. **c**, Calculated critical current, $I_{c,qd}$, as a function of gate voltage, V_{gate} , for $\delta/E_c = 2$. For odd numbers of electrons (red dots), the critical current is typically negative, similar to the measurement shown in Fig. 2a. **d**, $I_{c,qd}(V_{gate})$ for $\delta/E_c = 0.4$. Negative supercurrents are found for both odd (red dot) and even numbers of electrons (blue dot), as in the experimental data shown in **a**.

phase difference of π and therefore this event contributes to a negative supercurrent⁷ (see Supplementary Information). So, for a multi-level dot two effects can result in supercurrent reversal: permutation of tunnel events and an opposite parity of wavefunctions. When co-tunnelling events with a negative contribution dominate, the junction will exhibit a negative supercurrent. We note that the presence of the Kondo effect^{26,27} can result in a positive supercurrent where otherwise a negative supercurrent would be expected^{28,29}. We have not found any evidence for the Kondo effect in the normal state and therefore we disregard Kondo correlations in the modelling.

To further investigate the importance of multi-level effects, we numerically evaluate the critical current using fourth-order perturbation theory^{6,7} (see Supplementary Information for details). We assume that tunnel couplings are random in amplitude and sign (reflecting the parity of wavefunctions) and set $\Delta^*/E_c = 0.1$, as in our experiment. The probability for a negative supercurrent in the centre of a Coulomb diamond, P_π , is plotted in Fig. 4b for odd and even numbers of electrons. A very large average level spacing ($\delta/E_c \gg 1$) effectively gives a single-level quantum dot, so that $P_\pi = 1$ (0) for odd (even) numbers, as explained in Fig. 3a, b. The dependence of the critical current, $I_{c,qd}$, on V_{gate} (Fig. 4c) indeed unambiguously demonstrates the correlation between the number of electrons on the dot and the supercurrent sign. This correlation is absent in the opposite limit ($\delta/E_c \ll 1$), where $P_\pi \approx 0.3$ for both odd and even numbers of electrons, in agreement with previous calculations⁷. From the experimental data in Fig. 1e, we estimate $\delta/E_c \approx 0.4$, which clearly indicates an intermediate regime. Figure 4d shows a typical result for $I_{c,qd}$ versus V_{gate} for $\delta/E_c = 0.4$. As observed in the experiment, we obtain a negative supercurrent for both even (blue dot) and odd (red dot) numbers of electrons. Also, the typical line-shapes closely resemble the experimental data. Thus, in this multi-level regime, co-tunnelling events occur through a single level as well as through different levels. Consequently, the sign of the supercurrent is not only determined by the number of electrons on the quantum dot but also by the wavefunctions of the energy levels.

Received 25 April; accepted 22 June 2006.

- Josephson, B. D. Possible new effects in superconductive tunnelling. *Phys. Lett.* **1**, 251–253 (1962).
- Tinkham, M. *Introduction to Superconductivity* 2nd edn (McGraw-Hill, Singapore, 1996).
- Bulaevskii, L. N., Kuzii, V. V. & Sobyanin, A. A. Superconducting system with weak coupling to the current in the ground state. *JETP Lett.* **25**, 290–294 (1977).
- Glazman, L. I. & Matveev, K. A. Resonant Josephson current through Kondo impurities in a tunnel barrier. *JETP Lett.* **49**, 659–662 (1989).
- Spivak, B. I. & Kivelson, S. A. Negative local superfluid densities: The difference between dirty superconductors and dirty Bose liquids. *Phys. Rev. B* **43**, 3740–3743 (1991).
- Bauernschmitt, R., Siewert, J., Nazarov, Yu. V. & Odintsov, A. A. Josephson effect in low-capacitance superconductor–normal-metal–superconductor systems. *Phys. Rev. B* **49**, 4076–4081 (1994).
- Rozhkov, A. V., Arovas, D. P. & Guinea, F. Josephson coupling through a quantum dot. *Phys. Rev. B* **64**, 233301 (2001).
- van Harlingen, D. J. Phase-sensitive tests of the symmetry of the pairing state in the high-temperature superconductors - Evidence for $d_{x^2-y^2}$ symmetry. *Rev. Mod. Phys.* **67**, 515–535 (1995).
- Ryazanov, V. V. et al. Coupling of two superconductors through a ferromagnet: Evidence for a π -junction. *Phys. Rev. Lett.* **86**, 2427–2430 (2001).
- Baselmans, J. J. A., Morpurgo, A. F., van Wees, B. J. & Klapwijk, T. M. Reversing the direction of the supercurrent in a controllable Josephson junction. *Nature* **397**, 43–45 (1999).
- Ralph, D. C., Black, C. T. & Tinkham, M. Spectroscopic measurements of discrete electronic states in single metal particles. *Phys. Rev. Lett.* **74**, 3241–3244 (1995).
- Black, C. T., Ralph, D. C. & Tinkham, M. Spectroscopy of the superconducting gap in individual nanometer-scale aluminum particles. *Phys. Rev. Lett.* **76**, 688–691 (1996).
- Buitelaar, M. R., Nussbaumer, T. & Schönenberger, C. Quantum dot in the Kondo regime coupled to superconductors. *Phys. Rev. Lett.* **89**, 256801 (2002).
- Buitelaar, M. R. et al. Multiple Andreev reflections in a carbon nanotube quantum dot. *Phys. Rev. Lett.* **91**, 057005 (2003).
- Jarillo-Herrero, P., van Dam, J. A. & Kouwenhoven, L. P. Quantum supercurrent transistors in carbon nanotubes. *Nature* **439**, 953–956 (2006).
- Doh, Y. J. et al. Tunable supercurrent through semiconductor nanowires. *Science* **309**, 272–275 (2005).
- Wagner, R. S. & Ellis, W. C. Vapor-liquid-solid mechanism of single crystal growth. *Appl. Phys. Lett.* **4**, 89–90 (1964).
- Morales, A. M. & Lieber, C. M. A laser ablation method for the synthesis of crystalline semiconductor nanowires. *Science* **279**, 208–211 (1998).
- Björk, M. T. et al. One-dimensional heterostructures in semiconductor nanowhiskers. *Appl. Phys. Lett.* **80**, 1058–1060 (2002).
- Verheijen, M. A., Immink, G., de Smet, T., Borgström, M. T. & Bakkers, E. P. A. M. Growth kinetics of heterostructured GaP-GaAs nanowires. *J. Am. Chem. Soc.* **128**, 1353–1359 (2006).
- Sohn, L. L., Kouwenhoven, L. P. & Schön, G. (eds) *Mesoscopic Electron Transport* (Kluwer, Dordrecht, 1997).
- Baselmans, J. J. A., Heikkilä, T. T., van Wees, B. J. & Klapwijk, T. M. Direct observation of the transition from the conventional superconducting state to the π state in a controllable Josephson junction. *Phys. Rev. Lett.* **89**, 207002 (2002).
- Averin, D. V. & Nazarov, Y. V. in *Single Charge Tunneling* (eds Grabert, H. & Devoret, M. H.) Proc. NATO ASI Ser. B 294 217–247 (Plenum, New York, 1991).
- Björk, M. T. et al. Tunable effective g factor in InAs nanowire quantum dots. *Phys. Rev. B* **72**, 201307 (2005).
- Shimizu, Y., Horii, H., Takane, Y. & Isawa, Y. Multilevel effect on the Josephson current through a quantum dot. *J. Phys. Soc. Jpn* **67**, 1525–1528 (1998).
- Goldhaber-Gordon, D. et al. Kondo effect in a single-electron transistor. *Nature* **391**, 156–159 (1998).
- Cronenwett, S. M., Oosterkamp, T. H. & Kouwenhoven, L. P. A tunable Kondo effect in quantum dots. *Science* **281**, 540–544 (1998).
- Choi, M. S., Lee, M., Kang, K. & Belzig, W. Kondo effect and Josephson current through a quantum dot between two superconductors. *Phys. Rev. B* **70**, 020502 (2004).
- Siano, F. & Egger, R. Josephson current through a nanoscale magnetic quantum dot. *Phys. Rev. Lett.* **93**, 047002 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y.-J. Doh and L. Glazman for discussions, G. Immink for nanowire growth, and A. van der Enden and R. Schouten for technical support. Financial support was obtained from the Dutch Foundation for Fundamental Research on Matter (FOM), the Dutch Organisation for Scientific Research (NWO), the EU programmes HYSWITCH and NODE, and the Japanese International Cooperative Research Project (ICORP).

Author Contributions J.A.v.D., S.D.F. and L.P.K. are responsible for quantum transport experiments, Y.V.N. for numerical simulations, and E.P.A.M.B. for nanowire growth.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.P.K. (L.P.Kouwenhoven@TUDelft.nl).

Arctic hydrology during global warming at the Palaeocene/Eocene thermal maximum

Mark Pagani^{1*}, Nikolai Pedentchouk^{1*}, Matthew Huber^{2*}, Appy Sluijs³, Stefan Schouten⁴, Henk Brinkhuis³, Jaap S. Sinninghe Damsté^{4,5}, Gerald R. Dickens⁶ & the Expedition 302 Scientists†

The Palaeocene/Eocene thermal maximum represents a period of rapid, extreme global warming ~55 million years ago, superimposed on an already warm world^{1–3}. This warming is associated with a severe shoaling of the ocean calcite compensation depth⁴ and a >2.5 per mil negative carbon isotope excursion in marine and soil carbonates^{1–4}. Together these observations indicate a massive release of ¹³C-depleted carbon⁴ and greenhouse-gas-induced warming. Recently, sediments were recovered from the central Arctic Ocean⁵, providing the first opportunity to evaluate the environmental response at the North Pole at this time. Here we present stable hydrogen and carbon isotope measurements of terrestrial-plant- and aquatic-derived *n*-alkanes that record changes in hydrology, including surface water salinity and precipitation, and the global carbon cycle. Hydrogen isotope records are interpreted as documenting decreased rainout during moisture transport from lower latitudes and increased moisture delivery to the Arctic at the onset of the Palaeocene/Eocene thermal maximum, consistent with predictions of poleward storm track migrations during global warming⁶. The terrestrial-plant carbon isotope excursion (about –4.5 to –6 per mil) is substantially larger than those of marine carbonates. Previously, this offset was explained by the physiological response of plants to increases in surface humidity². But this mechanism is not an effective explanation in this wet Arctic setting, leading us to hypothesize that the true magnitude of the excursion—and associated carbon input—was greater than originally surmised. Greater carbon release and strong hydrological cycle feedbacks may help explain the maintenance of this unprecedented warmth.

PETM sediments, recovered from the Lomonosov ridge in the central Arctic Ocean during Integrated Ocean Drilling Program Expedition 302, are devoid of primary carbonates commonly used to assess palaeoclimate conditions. But they contain abundant organic components^{5,7}, including long-carbon-chain *n*-alkanes in the range *n*-C₂₃ to *n*-C₃₅, with a strong odd-over-even carbon-chain character, and short-chain *n*-alkanes dominated by *n*-C₁₇. Long-chain *n*-alkanes with an odd-over-even predominance typically derive from the waxes of higher plants⁸. However, this common interpretation was questioned in a single study of modern Arctic organic carbon from the Laptev Sea, which inferred a marine source for these compounds⁹. Importantly, this interpretation does not apply to the Arctic during the Palaeogene, where the evidence used to support a marine origin for long-chain *n*-alkanes is lacking (Supplementary Information). Short-chain *n*-alkanes represent a

mixture of aquatic sources, while *n*-alkane homologues dominated by *n*-C₁₇ typify algal and photosynthetic bacterial input¹⁰. Even-chain-length *n*-alkanes in the range *n*-C₁₄ to *n*-C₂₂ are ascribed to bacteria¹¹. Given lower abundances for even-chained *n*-alkanes during the PETM interval, we assume that *n*-C₁₇ derives predominantly from algal sources, in accordance with other studies^{12,13}.

To assess changes in the hydrological system during the PETM, we measured the stable hydrogen isotope (δD) values of *n*-C₂₉, *n*-C₂₇ and *n*-C₁₇ (Fig. 1b). Although our understanding of compound-specific hydrogen isotope systematics is incomplete, it is clear that source water hydrogen is a primary signal recorded by *n*-alkyl lipids^{14–16}. Our results implicate different water sources for high- and low-molecular-weight *n*-alkanes, with δD values of *n*-C₁₇ recording the δD of Arctic surface waters, and δD values of *n*-C₂₇ and *n*-C₂₉ (*n*-C_{27/29}) reflecting the hydrogen isotopic composition of precipitation in the Arctic region.

Data from contemporary environments suggest an apparent hydrogen isotope fractionation of –130‰ to –100‰ for higher-plant-derived *n*-C₂₉ and source water (Δ_{water-*n*-C₂₉})^{13,15,17}. Smaller Δ_{water-*n*-C₂₉} values (that is, less negative than –100‰) have been reported for deciduous conifers grown under continuous light¹⁸, similar to daylight conditions that characterize the Arctic growing season. A portion of this fractionation is potentially attributable to evapotranspiration, which acts to increase the δD value of leaf water used for lipid biosynthesis. However, rates of evapotranspiration are proportional to humidity¹⁶, and given evidence for both high humidity (Supplementary Information) and fluvial runoff⁷ during this interval of time, we assume that changes in the rates of evapotranspiration were negligible. Using Δ_{water-*n*-C₂₉} values of –130‰ and –100‰, we estimate that spring/summer precipitation had average δD values between –125‰ and –95‰ before the PETM, and –145‰ to –105‰ after the PETM, with much higher values (about –30‰ to –65‰) during the early phase of the event (Supplementary Information).

Our results suggest that Arctic PETM precipitation was substantially D-enriched relative to today, with hydrogen isotopic compositions comparable with modern, mid-latitude precipitation¹⁹. In general, the primary sources of atmospheric water vapour derive from the tropical and subtropical ocean. Poleward and altitudinal advection of air parcels approximately along isentropic surfaces²⁰ leads to cooling, condensation, an increase in the isotopic fractionation between the vapour and the condensate, and progressive isotopic distillation resulting in D-depleted high-latitude precipitation.

¹Department of Geology and Geophysics, Yale University, PO Box 208109, New Haven, Connecticut 06520, USA. ²Earth and Atmospheric Sciences Department and the Purdue Climate Change Research Center, Purdue University, 550 Stadium Mall Drive, West Lafayette, Indiana 47906, USA. ³Palaeoecology, Institute of Environmental Biology, Utrecht University, Laboratory of Palaeobotany and Palynology, Budapestlaan 4, 3584 CD, Utrecht, The Netherlands. ⁴Royal Netherlands Institute for Sea Research (NIOZ), Department of Marine Biogeochemistry and Toxicology, PO Box 59, 1790 AB, Den Burg, Texel, The Netherlands. ⁵Faculty of Geosciences, Department of Earth Sciences, Utrecht University, Budapestlaan 4, 3584 CD, Utrecht, The Netherlands. ⁶Department of Earth Sciences, Rice University, 6100 Main Street, Houston, Texas 77005, USA.

*These authors contributed equally to this work.

†A list of authors and affiliations appears at the end of the paper.

Deuterium-enriched Arctic precipitation during the PETM could have resulted from two end-member processes, including changes in proximal evaporative sources, or a decrease in large-scale (hemispheric to global) temperature gradients, such as the meridional temperature gradient as often assumed²¹, or the vertical temperature gradient along isentropes characteristic of baroclinic eddy-induced mixing. Changes in evaporative sources during the PETM, large enough to explain the observed δD shift, would require a fundamental, global alteration in precipitation and evaporation. However, such effects are not supported by physical modelling or proxy data (Supplementary Information) and thus not considered likely.

Alternatively, decreased meridional and/or vertical temperature gradients would conspire to reduce rainout of subtropical water vapour by synoptic eddies²², decrease isotopic distillation during vapour transport, and lead to D-enriched precipitation at high latitudes. If surface temperature gradients remained constant during the PETM⁷, then changes in atmospheric static stability could have decreased rainout in the mid-latitudes. In either case, a decrease in temperature gradients should be expressed as a reduction in the meridional isotopic gradient. Support for this supposition comes from soil carbonate $\delta^{18}O$ records, which suggest that the δD value of mid-latitude PETM precipitation increased by only $\sim 16\text{‰}$ (refs 23, 24), compared to a δD increase of $\sim 55\text{‰}$ in the Arctic (Fig. 1b).

As a reduced meridional isotopic gradient implies less rainout along the source airmass's trajectory, more water vapour must have been transported to extreme high latitudes. Observations that support an increase in water supply to the Arctic during the PETM include the prevalence of low-salinity-tolerant organic-walled dinoflagellate cyst (dinocyst) assemblages, evidence for photic zone anoxia indicative of low-salinity surface waters (Fig. 1c), and strong seasonal runoff⁷. This scenario suggests that the latitude of maximum

latent heat flux divergence (today at $\sim 40^\circ$) could have been situated much closer to the poles, broadly consistent with climate model results^{6,22}.

Enhanced moisture and latent heat transport from the subtropics to the Arctic region could have resulted from the nonlinear dependence of the saturation specific humidity of subtropical air parcels as a function of temperature^{1,6}, and/or a reduction of mid-latitude precipitation. Thus, as a corollary to our argument for an increase in poleward water vapour transport, we suggest that the subtropics and parts of the mid-latitudes experienced less net precipitation during the PETM.

Changes in surface water salinity were evaluated from δD_{nC17} . Limited analysis of lake sediments and aquatic plants suggest that values of $\Delta_{\text{water}-nC17}$ range from about -160‰ to -80‰ (refs 13, 15, 17). In order to determine plausible values of $\Delta_{\text{water}-nC17}$ for these Arctic sediments, we modelled the relationship between the δD of Arctic surface water, and the isotopic compositions of precipitation (that is, runoff) that acts to freshen the Arctic and alter its salinity (Fig. 2a). Our results suggest that δD values of Arctic surface waters were approximately -41‰ to -52‰ , resulting in a $\Delta_{\text{water}-nC17}$ of -70‰ to -80‰ .

The δD_{nC17} record, in conjunction with a $\Delta_{\text{water}-nC17}$ of -75‰ , indicate that the surface water salinity of the Arctic Ocean decreased as the PETM progressed, followed by a rapid increase in salinity towards the end of the climate anomaly (Fig. 2b). Low-salinity surface waters correspond with the presence of low-salinity-tolerant dinocyst assemblages and the occurrence of isorenieratene derivatives⁷ (Fig. 1c), biomarkers indicative of anoxia within the photic zone. This confluence suggests that water column stratification was promoted, in large part, by low-salinity surface water⁷, given that seawater density variations are determined primarily by salinity variations in the temperature range indicated by TEX_{86} temperature estimates (Fig. 1c). Termination of these conditions and a progressive

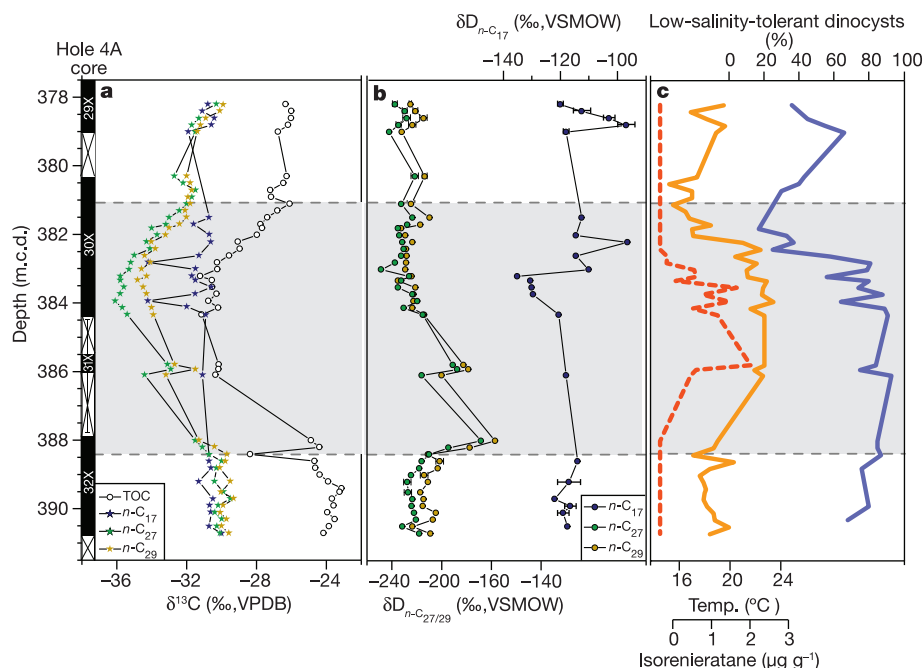


Figure 1 | Stable isotope results, and data on sea surface temperature, dinocysts and biomarkers. **a**, Stable carbon isotope data. m.c.d., metres composite depth below sea floor; TOC, total organic carbon. **b**, Stable hydrogen isotope data. Error bars represent the range of values based on duplicate measurements. **c**, Data from ref. 7. Orange line, sea surface temperatures based on the TEX_{86} index. Blue line, relative abundances of dinocysts produced by dinoflagellate species that were tolerant of low

salinity. Red dashed line, absolute abundances of isorenieratene derivatives, biomarkers derived from photosynthetic green sulphur bacteria. The presence of these bacteria implies water column stratification and the development of photic zone euxinia. The leftmost (recovery) column shows core numbers: black bars, recovered material; white bars, missing material. Error bars connected to Core 31X indicate the uncertainty of its stratigraphic position (see ref. 7). Grey-shaded area indicates the PETM.

change to fewer low-salinity-tolerant dinocyst assemblages⁷ coincide with a trend towards more saline surface waters by the end of the climate anomaly (Fig. 2b).

Further insight into the nature of Arctic climate change is provided by the carbon isotopic ($\delta^{13}\text{C}$) compositions of the same *n*-alkyl lipids. The $\delta^{13}\text{C}$ values of *n*-C_{29/27} show a large negative carbon isotope excursion (CIE) across the PETM that tracks the $\delta^{13}\text{C}$ record of total organic carbon⁷ ($\delta^{13}\text{C}_{\text{TOC}}$) (Fig. 1a), but are $\sim 6\text{‰}$ more negative than $\delta^{13}\text{C}_{\text{TOC}}$ values before and after the CIE, and $\sim 3\text{‰}$ to 5‰ more negative during the CIE. This correlation, both in trend and isotopic offset, suggests that the TOC is substantially influenced by terrestrial components.

Importantly, the CIE from higher plant *n*-alkanes (approximately -4.5‰ to -6‰) is substantially larger than the CIE anomaly (about -2.5‰ to -3‰) generally recorded in bulk marine carbonate and benthic foraminifera^{1,3}, but similar to those recorded in soil carbonates from mid-latitudes² and some planktonic foraminiferal records³. To explain the larger terrestrial CIE elsewhere, it has been suggested that soil moisture and humidity increased during the PETM, amplifying the carbon isotopic fractionation associated with terrestrial photosynthesis (CIF)². In this model, a larger CIF due to increasing humidity and soil moisture has to overcome the antagonistic effect of decreasing CIF due to higher temperatures during the PETM^{2,7}. In order for this model to explain our data, latest Palaeocene soil moistures and humidity levels surrounding the Arctic Ocean would have to have been quite low, with soil moistures at or below 40% (ref. 2), similar to those found in modern arid-to-semiarid environments (Supplementary Information). A dry Arctic is difficult to justify, given palaeoecological reconstructions from high-latitude coal distributions and terrestrial palaeobotanical records that strongly support the prevalence of mesic to wet climate conditions throughout the Arctic region during the Palaeocene and Eocene (Supplementary Information). Conditions associated with analogous modern tropical and warm temperature forests can be used to predict that the Palaeogene Arctic maintained an average relative humidity of $>70\%$, with soil moistures $>60\%$ (Supplementary Information)—estimates that agree well with fully coupled climate model simulations (Supplementary Information). The accumulative

evidence, including the δD results from this study, and evidence for high seasonal discharge and low-salinity surface waters during the PETM⁷, point to a warm and moist late Palaeocene–early Eocene Arctic region, with increased precipitation expressing itself largely as increased runoff⁷.

Other possibilities exist to explain this CIE discrepancy between higher plant *n*-alkanes and marine carbonate records. For example, given potentially more negative $\delta^{13}\text{C}$ values for angiosperms relative to gymnosperms²⁵, an increase in angiosperms relative to gymnosperms during the PETM⁷ could possibly explain a larger isotopic excursion in $\delta^{13}\text{C}_{\text{TOC}}$ relative to marine carbonates. However, modern isotopic offsets between gymnosperms and angiosperms probably result from differences in water use efficiency²⁵ under low CO_2 concentrations. Therefore, more humid environmental conditions during the PETM in the Arctic would have arguably minimized isotopic distinctions between these two plant types. In addition, the $\delta^{13}\text{C}$ difference between *n*-alkanes and bulk leaf carbon is potentially smaller for angiosperms relative to gymnosperms¹⁷. Thus, the $\delta^{13}\text{C}$ excursion in *n*-C_{29/27} potentially represents a minimum change—if flora shifts occurred in the Arctic region during the PETM, and whole plant $\delta^{13}\text{C}$ values of gymnosperms and angiosperms were similar. Finally, similar ecosystem shifts would need to have synchronously occurred globally to account for terrestrial ^{13}C excursions of the same magnitude elsewhere.

Rather than systematically explaining a similar observation by different processes, it is possible that the CIE expressed by higher plant *n*-alkanes reflects the true $\delta^{13}\text{C}$ change of atmospheric carbon dioxide in equilibrium with the ocean during the PETM. Such a scenario implies that foraminiferal $\delta^{13}\text{C}$ values do not accurately represent the full CIE of dissolved inorganic carbon in the ocean owing to effects related to dissolution⁴ and changes in pH (ref. 26). We note that in marine cores with limited carbonate dissolution, the observed $\delta^{13}\text{C}$ excursion recorded from shallow-dwelling plankton is about -4‰ (ref. 3), which is close to our value. Further consideration of the carbon isotope effect related to changes in pH allow for the ^{13}C -enrichment of carbonates in the range of 0.5‰ (refs 1, 2). Therefore, it is possible to account for rather similar terrestrial and marine CIEs. Importantly, evidence for a CIE of about -4.5‰ to -5‰ nearly doubles the mass estimate for the release of carbon during this time, and more closely approximates the carbon concentration required to account for the observed shoaling of the ocean carbonate compensation depth⁴, as well as changes in global temperatures.

The $\delta^{13}\text{C}$ values of *n*-C₁₇ are nominally -30‰ , similar to those of *n*-C_{29/27} before and following the PETM, but are ^{13}C -enriched relative to higher plant *n*-alkanes during the PETM (Fig. 1a). In modern settings, algal organic carbon is consistently ^{13}C -enriched relative to terrestrial C₃ plants. Similar $\delta^{13}\text{C}$ values for algal and higher plant organic carbon before and after the PETM is consistent with a high atmospheric CO_2 environment^{27,28}. However, the relative increase in the $\delta^{13}\text{C}$ of *n*-C₁₇ during the PETM must have resulted from physiological factors that overwhelmed the influence of ^{13}C -depleted carbon, such as increases in the volume to surface area of cells, and/or algal growth rates²⁹, consistent with evidence for eutrophy and nutrient-rich conditions during the PETM⁷. Interestingly, changes in primary productivity appear to have occurred in other ocean regions³⁰ and are thought to have played a critical role in the carbon cycle response to environmental change. Given the relative isolation of the Arctic from other ocean basins during this time⁷, widespread changes in primary production were probably linked to changes in riverine nutrient supply; this implicates the hydrological cycle as an important agent forcing biological and environmental change during the PETM.

METHODS

Sample extraction. Sediments were extracted with dichloromethane using an accelerated solvent extraction system (ASE 300; Dionex Corporation) at 125°C ,

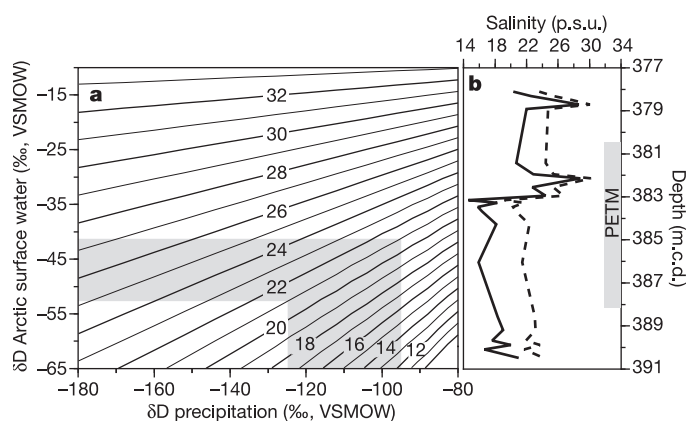


Figure 2 | δD of precipitation and Arctic surface water, and salinity. **a**, Calculations³¹ of Arctic Ocean salinity (numbers on contours, in practical salinity units, p.s.u.) assume an ice-free, global seawater δD value of -8‰ (VSMOW) and a global salinity of 34 p.s.u. Shaded bands represent pre-PETM conditions, given a salinity of 21 p.s.u. (Supplementary Information) and δD values of precipitation from $\delta\text{D}_{\text{H}_2\text{O}}$, yielding a δD estimate for Arctic surface water ($\delta\text{D}_{\text{AW}}$) of about -41‰ to -52‰ , resulting in a $\Delta_{\text{water-nC}_{17}}$ value (see text) of about -70‰ to -80‰ . **b**, Surface water salinity calculations, derived from $\delta\text{D}_{\text{H}_2\text{O}}$; $\delta\text{D}_{\text{AW}}$ was calculated from $\delta\text{D}_{\text{H}_2\text{O}}$, assuming a $\Delta_{\text{water-nC}_{17}}$ value of -75‰ , and the δD of precipitation derived from $\delta\text{D}_{\text{H}_2\text{O}}$ assuming $\Delta_{\text{water-nC}_{29}}$ values of -100‰ (solid line) and -130‰ (dashed line).

1,500 p.s.i., for 25 min. Lipid fractions were separated by column chromatography (70–230 mesh) using an elution sequence of hexane, hexane/dichloromethane (9:1 v/v) and dichloromethane/methanol (2:1 v/v). Cyclic and branched alkanes were separated from normal and isoalkanes by adduction with urea. The hydrocarbon fraction was dried under a stream of N_2 and dissolved in a mixture of methanol-saturated urea, pentane and acetone (200 μ l each). The resulting urea crystals were extracted with hexane, yielding cyclic/branched alkanes. Remaining urea crystals were dissolved in 500 μ l of H_2O and 500 μ l of methanol, then extracted with hexane to yield the *n*-alkane fraction.

Analysis. The adducted fraction was analysed for stable carbon and hydrogen isotopic compositions on a Thermo Finnigan MAT 253 mass spectrometer interfaced with a Thermo Finnigan Trace GC Combustion III (for carbon measurements) and High Temperature Conversion (for hydrogen measurements) systems. Individual *n*-alkanes were separated using a J&W Scientific DB-1 capillary column (60 m $\times$ 0.25 mm $\times$ 0.25 μ m). The gas chromatograph oven was programmed from 60 $^{\circ}C$ (held for 1 min) at 6 $^{\circ}C$ min $^{-1}$ to 320 $^{\circ}C$ and held for 25 min isothermally. A programmed temperature vaporizing injector was used. Helium was used as a carrier gas with a column flow rate of 2.0 ml min $^{-1}$. Carbon isotopic compositions are expressed relative to the VPDB standard, based on an in-house reference gas calibrated against the OzTech standard ($\delta^{13}C = -40.61\text{‰}$). The analytical accuracy and precision of the MAT 253 mass spectrometer during carbon isotope measurements were $\pm 0.2\text{‰}$ (the root-mean-square error), based on an *n*-C₂₀ alkane standard injected daily. The standard error of *n*-alkane $\delta^{13}C$ measurements (based on duplicates of 22 samples) was $\pm 0.8\text{‰}$ or better (Supplementary Information).

Pyrolytic conversion of organic hydrogen to H_2 was conducted at 1,400 $^{\circ}C$. Measurements of the H_3^+ factor (the proportionality constant between the concentration of H_2 and H_3^+ in the mass spectrometer) were determined daily using H_2 reference gas. The H_3^+ factor varied between 15.27 and 16.11 p.p.m. nA $^{-1}$ over a period of 23 days, averaging 15.70 p.p.m. nA $^{-1}$, with a standard deviation of 0.24 p.p.m. nA $^{-1}$. H_2 peak heights varied over an eightfold range, which was in the range of most of the H_2 peaks from analysed compounds.

Errors. The precision of isotopic measurements of H_2 reference gas after H_3^+ correction was $\pm 0.9\text{‰}$ or better. The analytical accuracy and precision of the system were determined using an externally co-injected standard mixture of *n*-C₁₆ to *n*-C₃₀ alkanes and 5 α -androstane (isotopic ratios were measured off-line by A. Schimmelmann, Biogeochemical Laboratories, Indiana University), which were analysed at least once per day. The root-mean-square error for hydrogen isotopic measurements of these compounds was 3.4 ‰ ($n = 240$). Hydrogen isotopic compositions of *n*-alkanes are reported based on duplicate analyses. δD values are expressed relative to the VSMOW standard, based on an in-house reference gas adjusted daily using a 5 α -androstane standard. The standard error of *n*-C₁₇, *n*-C₂₇ and *n*-C₂₉ δD measurements was generally less than $\pm 5\text{‰}$, reaching $\pm 6\text{‰}$ only in a few cases (see Supplementary Information).

Received 6 July 2005; accepted 3 July 2006.

1. Zachos, J. *et al.* A transient rise in tropical sea surface temperature during the Paleocene-Eocene thermal maximum. *Science* **302**, 1551–1554 (2003).
2. Bowen, G. J., Beerling, D. J., Koch, P. L., Zachos, J. C. & Quattlebaum, T. A humid climate state during the Palaeocene/Eocene thermal maximum. *Nature* **432**, 495–499 (2004).
3. Thomas, D. J., Zachos, J. C., Bralower, T. J., Thomas, E. & Bohaty, S. Warming the fuel for the fire: evidence for the thermal dissociation of methane hydrate during the Paleocene-Eocene thermal maximum. *Geology* **30**, 1067–1070 (2002).
4. Zachos, J. C. *et al.* Rapid acidification of the ocean during the Paleocene-Eocene thermal maximum. *Science* **308**, 1611–1615 (2005).
5. Backman, J., Moran, K., McInroy, D. B., Mayer, L. A. & Expedition Scientists. Arctic Coring Expedition (ACEX). *Proc. IODP* **302** 1/doi:10.2204/iodp.proc.302.2006 (Integrated Ocean Drilling Program Management International, College Station, Texas, 2006).
6. Caballero, R. & Langen, P. The dynamic range of poleward energy transport in an atmospheric general circulation model. *Geophys. Res. Lett.* **32**, doi:10.1029/2004GL021581 (2005).
7. Sluijs, A. *et al.* Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum. *Nature* **441**, 610–613 (2006).
8. Eglinton, G. & Hamilton, R. J. Leaf epicuticular waxes. *Science* **156**, 1322–1335 (1967).
9. Zegouagh, Y., Derenne, S., Largeau, C., Bardoux, G. & Mariotti, A. Organic matter sources and early diagenetic alteration in Arctic surface sediments (Lena River delta and Laptev Sea, Eastern Siberia), II. Molecular and isotopic studies of hydrocarbons. *Org. Geochem.* **28**, 571–583 (1998).
10. Han, J. & Calvin, M. Hydrocarbon distribution of algae and bacteria, and

microbiological activity in sediments. *Proc. Natl Acad. Sci. USA* **64**, 436–443 (1969).

11. Grimalt, J. & Albaiges, J. Sources and occurrence of C_{12–22} *n*-alkane distributions with even carbon-number preference in sedimentary environments. *Geochim. Cosmochim. Acta* **51**, 1379–1384 (1987).
12. Muri, G., Wakeham, S. G., Pease, T. K. & Faganeli, J. Evaluation of lipid biomarkers as indicators of changes in organic matter delivery to sediments from Lake Planina, a remote mountain lake in NW Slovenia. *Org. Geochem.* **35**, 1083–1093 (2004).
13. Sachse, D., Radke, J. & Gleixner, G. Hydrogen isotope ratios of recent lacustrine sedimentary *n*-alkanes record modern climate variability. *Geochim. Cosmochim. Acta* **68**, 4877–4889 (2004).
14. Sternberg, L. D. D/H ratios of environmental water recorded by D/H ratios of plant lipids. *Nature* **333**, 59–61 (1988).
15. Sauer, P. E., Eglinton, T. I., Hayes, J. M., Schimmelmann, A. & Sessions, A. L. Compound-specific D/H ratios of lipid biomarkers from sediments as a proxy for environmental and climatic conditions. *Geochim. Cosmochim. Acta* **65**, 213–222 (2001).
16. Yakir, D. in *Stable Isotopes* (ed. Griffiths, H.) 147–168 (BIOS Scientific Publishers, Oxford, 1998).
17. Chikaraishi, Y. & Naraoka, H. Compound-specific δD and $\delta^{13}C$ analyses of *n*-alkanes extracted from terrestrial and aquatic plants. *Phytochemistry* **63**, 361–371 (2003).
18. Yang, H., Equiza, A. M., Jagels, R., Pagani, M. & Briggs, D. Carbon and hydrogen isotopic compositions of deciduous conifers under a continuous-light environment: implications for the interpretation of the high-latitude plant isotope record at the PETM. (Salt Lake City Annual Meeting, October 16–19, The Geological Society of America, 2005).
19. Bowen, G. J. & Revenaugh, J. Interpolating the isotopic composition of modern meteoric precipitation. *Wat. Resour. Res.* **39**, doi:10.1029/2003WR002086 (2003).
20. Pierrehumbert, R. T. Lateral mixing as a source of subtropical water vapor. *Geophys. Res. Lett.* **25**, 151–154 (1998).
21. Boyle, E. A. Cool tropical temperatures shift the global $\delta^{18}O$ -T relationship: An explanation for the ice core $\delta^{18}O$ -borehole thermometry conflict? *Geophys. Res. Lett.* **24**, 273–276 (1997).
22. Pierrehumbert, R. T. The hydrologic cycle in deep time climate problems. *Nature* **419**, 191–198 (2002).
23. Koch, P. L. *et al.* in *Causes and Consequences of Globally Warm Climates in the Early Paleogene* (eds Wing, S. L., Gingerich, P. D., Schmitz, B. & Thomas, E.) 49–64 (Special Paper 369, Geological Society of America, Boulder, 2003).
24. Fricke, H. C., Clyde, W. C., O'Neil, J. R. & Gingerich, P. D. Evidence for rapid climate change in North America during the latest Paleocene thermal maximum: oxygen isotope compositions of biogenic phosphate from the Bighorn Basin (Wyoming). *Earth Planet. Sci. Lett.* **160**, 193–208 (1998).
25. Leavitt, S. W. & Newberry, T. Systematics of stable-carbon isotopic differences between gymnosperm and angiosperm trees. *Plant Physiol.* **11**, 257–262 (1992).
26. Spero, H. J., Bijma, J., Lea, D. W. & Bernis, B. E. Effect of seawater carbonate concentration on foraminiferal carbon and oxygen isotopes. *Nature* **390**, 497–500 (1997).
27. Arthur, M. A., Walter, D. E. & Claypool, G. E. Anomalous ^{13}C enrichment in modern marine organic carbon. *Nature* **315**, 216–218 (1985).
28. Pagani, M., Zachos, J. C., Freeman, K. H., Tipler, B. & Bohaty, S. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309**, 600–603 (2005).
29. Popp, B. *et al.* Effect of phytoplankton cell geometry on carbon isotopic fractionation. *Geochim. Cosmochim. Acta* **62**, 69–77 (1998).
30. Bujak, J. P. & Brinkhuis, H. in *Late Paleocene - Early Eocene Biotic and Climatic Events in the Marine and Terrestrial Records* (eds Aubry, M.-P., Lucas, S. G. & Berggren, W. A.) 277–295 (Columbia Univ. Press, New York, 1998).
31. Railsback, L. B., Anderson, T. F., Ackerly, S. C. & Cisne, J. L. Paleocceanographic modeling of temperature-salinity profiles from stable isotopic data. *Paleoceanography* **4**, 585–591 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements M.P. thanks K. Turekian for conversations, and G. Bowen for comments and suggestions that substantially improved the manuscript. M.H. thanks the Purdue Research Foundation, ITaP, NCAR, NSF and L. C. Sloan for support for this research. A.S. thanks Utrecht Biogeology Centre for funding. H.B. thanks NWO, the Netherlands Organization for Scientific Research, and Utrecht University for enabling participation. We appreciate technical assistance provided by C. Valache, A. McLawhorn and G. Olack. This research used samples and data provided by the Integrated Ocean Drilling Program (IODP), which is sponsored by the US NSF and participating countries under the management of Joint Oceanographic Institutions (JOI) Inc.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.P. (mark.pagani@yale.edu).

Expedition 302 Scientists (those not listed above): Jan Backman¹, Steve Clemens², Thomas Cronin³, Frédérique Eynaud⁴, Jérôme Gattacceca⁵, Martin Jakobsson⁶, Ric Jordan⁷, Michael Kaminski⁸, John King⁹, Nalân Koc¹⁰, Nahysa C. Martinez¹¹, David McInroy¹², Theodore C. Moore Jr¹³, Matthew O'Regan⁹, Jonaotaro Onodera¹⁴, Heiko Pälke¹⁵, Brice Rea¹⁶, Domenico Rio¹⁷, Tatsuhiko Sakamoto¹⁸, David C. Smith⁹, Kristen E. K. St John¹⁹, Itsuki Suto²⁰, Noritoshi Suzuki²¹, Kozo Takahashi¹⁴, Mahito Watanabe²² & Masanobu Yamamoto²³

¹Department of Geology and Geochemistry, Stockholm University, Stockholm, SE-10691, Sweden. ²Geological Sciences, Brown University, 324 Brook Street, PO Box 1846, Providence, Rhode Island 02912-1846, USA. ³US Geological Survey, Eastern Earth Surface Processes Team, 926A USGS National Center, Reston, Virginia 20192, USA. ⁴Département de Géologie et Océanographie, Université Bordeaux 1, Avenue des facultés, c/o Bernei Housen, 33405 Talence Cedex, France. ⁵Department of Geophysics, CEREGE (CNRS)/University of Aix-Marseille 3, BP80, 13545 Aix-en-Provence Cedex 4, France. ⁶Department of Geology and Geochemistry, Stockholm University, 10691 Stockholm, Sweden. ⁷Department of Earth and Environmental Sciences, Faculty of Science, Yamagata University, 1-4-12 Kojirakawa-machi, Yamagata 990-8560, Japan. ⁸Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK. ⁹Graduate School of Oceanography, University of Rhode Island, Narragansett Bay Campus, South Ferry Road, Narragansett, Rhode Island 02882, USA. ¹⁰Norwegian Polar Institute, Polar Environmental Center, N-9296 Tromsø, Norway. ¹¹Department of Earth Sciences, Boston University, 685 Commonwealth Avenue, Boston, Massachusetts 02215, USA. ¹²British Geological Survey, Murchison House, West Mains Road, Edinburgh EH9 3LA, UK. ¹³Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109-1063, USA. ¹⁴Department of Earth and Planetary Sciences, Graduate School of Sciences, Kyushu University, Hakozaki 6-10-1, Higashi-ku, Fukuoka 812-8581, Japan. ¹⁵School of Ocean and Earth Science, University of Southampton, Southampton Oceanography Centre, European Way, Southampton SO14 3ZH, UK. ¹⁶Department of Geography and Environment, School of Geosciences, University of Aberdeen, Elphinstone Road, Aberdeen AB24 3UF, UK. ¹⁷Department of Geology, Paleontology and Geophysics, University of Padova, Via Giotto 1 I-35137 Padova, Italy. ¹⁸Institute for Research on Earth Evolution (IFREE), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Natsushima-cho 2-15, Yokosuka 237-0061, Japan. ¹⁹Department of Geology and Environmental Science, MSC7703, James Madison University, Harrisonburg, Virginia 22807, USA. ²⁰Institute of Life and Environmental Science, University of Tsukuba, Tennoudai 1-1-1, Tsukuba, Ibaraki 305-8572, Japan. ²¹Institute of Geology and Paleontology, Graduate School of Science, Tohoku University, Aramaki, Aoba, Aoba-ku, Sendai City 980-8578, Japan. ²²Institute of Geoscience, National Institute of Advanced Industrial Science, and Technology (Geological Survey of Japan), AIST Tsukuba Central 7, Higashi-1-1-1, Tsukuba, Ibaraki 305-8567, Japan. ²³Graduate School of Environmental Earth Science, Hokkaido University, Kita-10, Nishi-5, Kita-ku, Sapporo 060-0810, Japan.

LETTERS

The dynamics of melt and shear localization in partially molten aggregates

Richard F. Katz¹, Marc Spiegelman^{1,2} & Benjamin Holtzman¹

The volcanoes that lie along the Earth's tectonic boundaries are fed by melt generated in the mantle. How this melt is extracted and focused to the volcanoes, however, remains an unresolved question. Here we present new theoretical results with implications for melt focusing beneath mid-ocean ridges. By modelling laboratory experiments^{1,2}, we test a formulation for magma dynamics and provide an explanation for localized bands of high-porosity and concentrated shear deformation observed in experiments. These bands emerge and persist at 15°–25° to the plane of shear. Past theoretical work on this system predicted the emergence of melt bands^{3,4} but at an angle inconsistent with experiments. Our results suggest that the observed band angle results from a balance of porosity-weakening and strain-rate-weakening deformation mechanisms. Lower band angles are predicted for greater strain-rate weakening. From these lower band angles, we estimate the orientation of melt bands beneath mid-ocean ridges and show that they may enhance magma focusing toward the ridge axis.

Recent experiments^{1,2} demonstrate that partially molten aggregates deformed in simple shear develop localized melt bands of high porosity and enhanced strain (Fig. 1a). These bands emerge at low angles (~20°) to the plane of shear for a range of strain rates and stresses, and persist at low angles even after large shear strains. This pattern-forming instability presents a rare opportunity to test theories of magma transport in the Earth's mantle^{5–8}. Magma dynamics theories use continuum equations for conservation of mass, momentum and energy to describe a two-phase system of low-viscosity magma in a deformable, permeable solid matrix and should be applicable to the experiments. Past theoretical work³ showed that a porosity-weakening viscous material⁹ undergoing extension is unstable: tension across a weak, high-porosity region leads to low pressure that, in turn, causes convergence of melt flow into that region, raising its porosity and further weakening it. This instability has been predicted to occur at scales smaller than the compaction length^{3,4,10,11}, which is the intrinsic length-scale in magma dynamics theory⁵.

Past theoretical work predicts that melt bands emerge perpendicular to the direction of the maximum rate of extensional strain. This prediction results from assuming that the matrix viscosity depends only on porosity and weakens with increasing melt fraction. For simple shear geometry, Spiegelman⁴ showed that bands oriented at 45° to the shear plane will grow fastest, whereas melt bands with angles greater than 90° will decay (Supplementary Fig. S1). Here we demonstrate that a viscosity that includes both porosity and strain-rate-weakening mechanisms can reproduce the emergence and persistence of melt bands at about 20° to the direction of maximum shear (a difference of 25° from past predictions), as observed in experiments.

A power-law form for strain-rate weakening is a commonly

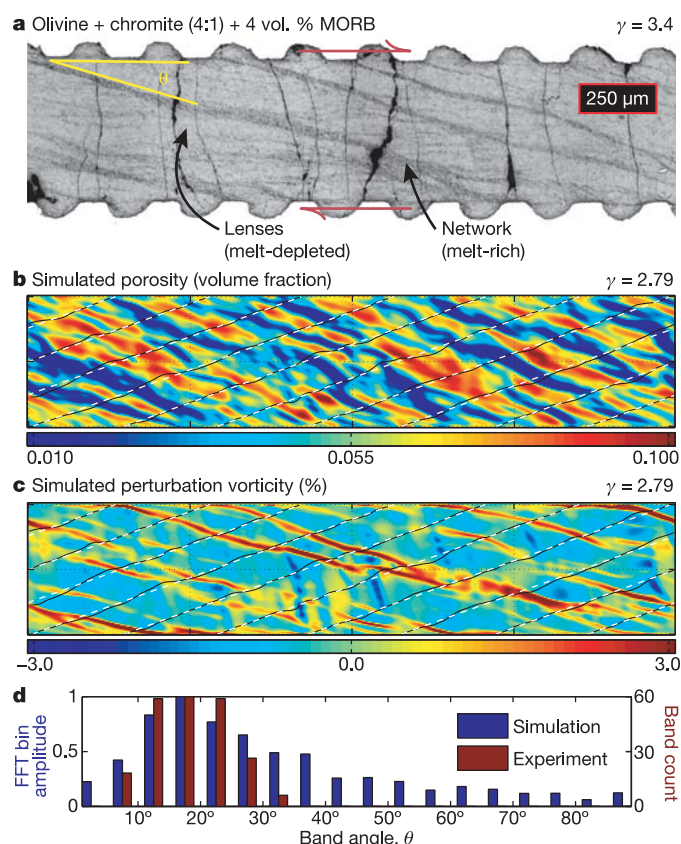


Figure 1 | A comparison of experimental and numerical results. **a**, An example cross-section of an experiment (PI-1096) on a partially molten olivine–basalt–chromite aggregate deformed in simple shear to a strain of 3.4. (Adapted from Fig. 1a of ref. 18; experimental details in ref. 2.) The melt-rich bands are sloping, darker-grey regions at an angle θ to the shear plane. Sub-vertical black features are decompression cracks, an experimental artefact. **b**, **c**, The porosity (**b**) and perturbation vorticity (**c**) from a numerical simulation with $n = 6$ and $\alpha = -27$ at a shear strain γ of 2.79. The domain is five by one compaction lengths, approximately equal to the estimated size of an experimental charge. The perturbation vorticity, $\nabla \times [V - \gamma y] / \dot{\gamma}$, is the total vorticity minus the constant vorticity $\dot{\gamma}$ due to simple shear (here normalized by $\dot{\gamma}$). Black lines in **b** and **c** show the position of passive tracer particles that were arrayed in vertical lines at $\gamma = 0$; white dotted lines show the expected position of the tracers due only to simple shear. The linear, low-angle red bands in **c** are weak regions associated with high porosity and enhanced shear, while the linear, sub-vertical blue regions are regions of reversed shear. **d**, Histograms comparing band-angle distributions in experiments and the numerical solution from **b**.

¹Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA. ²Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York 10027, USA.

accepted constitutive relation for high-temperature creep of mantle materials^{9,12}. Here we assume^{9,12–14}:

$$\eta(\phi, \dot{\epsilon}) = \eta_0 e^{\alpha(\phi - \phi_0)} \dot{\epsilon}_{II}^{\frac{1-n}{n}} \quad (1)$$

where η_0 is the shear viscosity at the reference porosity ϕ_0 and reference strain rate. $\alpha = -28 \pm 3$ is an experimentally derived porosity-weakening coefficient^{14,15}, $\dot{\epsilon}_{II}$ is the second invariant of the incompressible component of the strain-rate tensor, and n defines the power-law dependence of viscosity on stress. This viscosity is newtonian when $n = 1$ and is a standard non-newtonian power-law viscosity when $n > 1$ and $\phi = 0$.

To understand the effect of the strain-rate dependence of viscosity we extend a previous linear analysis⁴. Formally, we introduce an infinitesimal plane-wave perturbation (analogous to a melt-rich band) to a constant-porosity system undergoing simple shear and solve for its growth as a function of orientation (Supplementary Information). The growth rate, $\dot{s}$, of porosity bands predicted by this analysis is:

$$\dot{s}(\theta, n) = -\frac{\alpha \xi (1 - \phi_0) \sin(2\theta)}{1 + \frac{1-n}{n} \cos^2(2\theta)} \quad (2)$$

where $\theta(t)$ is the angle between the melt bands and the shear plane (Fig. 1a), which increases with time, t , due to advection by the shear flow. The parameter $\xi = (\xi_0/\eta_0 + 4/3)^{-1}$ depends on the ratio of bulk viscosity to shear viscosity and controls the growth rate of band

porosity through the product $\alpha \xi$. The amplitude of porosity in the bands at strain $\gamma(t)$ is given by $e^{s(t)}$.

Figure 2a shows the growth rate of melt bands as a function of their angle to the shear plane, θ , and the stress exponent, n . For a viscosity that weakens with porosity only, melt bands oriented at 45° grow fastest because they are perpendicular to the direction of maximum extension in simple shear (equation (2) with $n = 1$).

For a strain-rate-dependent viscosity ($n > 1$), however, concentrated shear deformation further weakens the bands, allowing them to more easily de-compact under extension. Enhanced shear strain is largest for porosity bands oriented at zero and 90° and goes to zero for 45° bands (Supplementary Fig. S1 and $\cos^2(2\theta)$ term in equation (2)). Thus two competing processes affect the preferred angle of melt bands. The balance between favourable orientation for extension (45°) and favourable orientation for concentrating shear (0 and 90°) is controlled by the factor $(1 - n)/n$. Figure 2a shows the effect of changing n on the growth rate of porosity. As n increases from 1 to 6, the single peak in $\dot{s}$ at 45° broadens and divides into symmetric peaks at low ($\sim 15^\circ$) and high ($\sim 75^\circ$) angle.

Although the instantaneous growth rate of melt bands in Fig. 2a is symmetric about 45° , advection by simple shear affects low- and high-angle bands differently. Low-angle bands are rotated slowly and persist in favourable orientations for growth of porosity. High-angle bands, however, are rapidly rotated away from angles with positive growth rates. Figure 2b shows that after a strain of $\gamma = 1$, bands at low

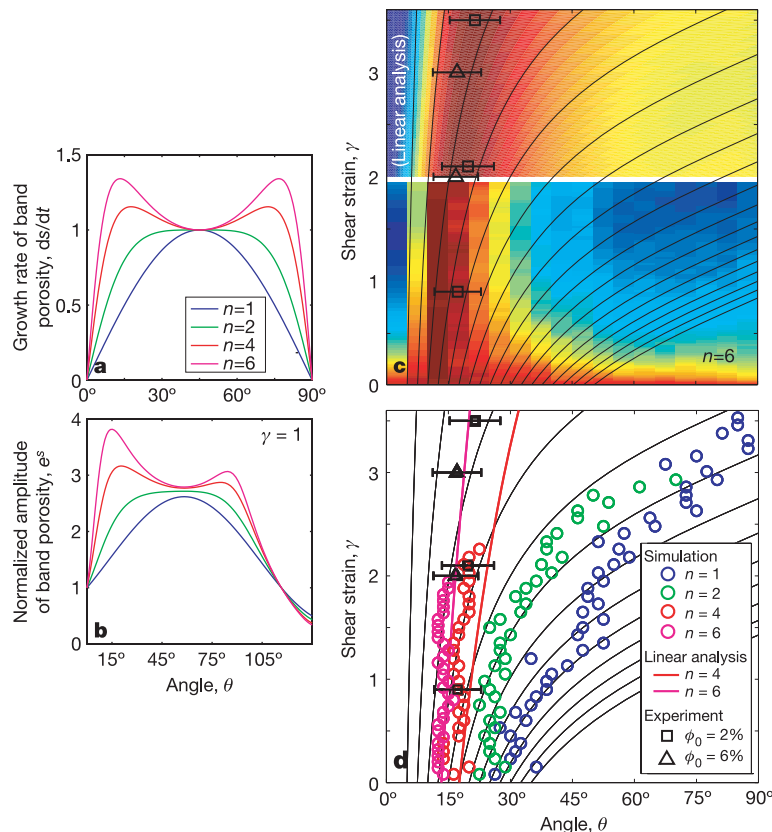


Figure 2 | Results from linear analysis and numerical simulations, showing the effect of stress exponent n on the angular dependence of porosity of melt bands. **a**, The growth rate $\dot{s}$ of porosity as a function of angle; equation (2). **b**, The normalized amplitude of melt-band porosity as a function of θ and n for $\gamma = 1$. Advection by simple shear preferentially rotates higher-angle bands out of favourable orientation for growth. The angle with peak amplitude at a given strain is determined by the growth rate and the rate of passive rotation by simple shear. **c**, Normalized, binned amplitude of FFTs of the porosity field from an ensemble of numerical simulations with $n = 6$ ($n = 1, 4$ shown in Supplementary Fig. S4). Linear analysis is used to extend simulation results to

$\gamma = 3.6$ (Supplementary Fig. S2 shows angle evolution from linear analysis only). The colour scale goes from zero to one. **d**, Summary of simulation ensemble results for $n = 1, 2, 4$ and 6 . Coloured symbols represent the peak of the band-angle histogram and each is consistent with the angle estimated by visual inspection of the porosity field. Black symbols with 1-sigma error bars represent mean band angles in experiments, quantified by hand-measurement of sectioned experimental-run products. Black lines demonstrate the rotational effect of unperturbed simple shear on selected initial band angles (Supplementary equation S17). The behaviour of full nonlinear calculations for high strain at $n > 3$ is still unclear.

angle have a larger amplitude than those at high angle. The systematics of band rotation are detailed in Supplementary equation S17.

The linear analysis suggests that increasing the contribution of strain-rate weakening favours the growth of low-angle melt bands. To validate these results and to explore the behaviour of the full non-linear system requires computational solutions, which we have performed using the Portable Extensible Toolkit for Scientific computation^{16,17} (see Supplementary Information). The simulations are initiated at a constant porosity plus <1% random noise. Output from a representative run is shown in Fig. 1b and c.

Histograms of the distribution of band angles are calculated from simulations by taking the two-dimensional fast fourier transform (FFT) of the porosity field and integrating the amplitude response within 5° bins in band angle θ (Fig. 1d). An ensemble of simulations is run, each with the same n but different initial noise, to produce a composite histogram at each time-step. Figure 2c shows the evolution of the band-angle distribution with increasing strain for $n = 6$. The peak of this histogram is initially $\sim 15^\circ$ and persists at low angle with increasing strain. The evolution of the peak angle is also shown by magenta circles in Fig. 2d, which charts the peak angle for ensembles of simulations with different n . Consistent with the linear analysis, a more nonlinear viscosity (higher n) results in a lower initial peak band angle.

In Fig. 2d, the angle corresponding to the peak of each of the histograms increases with strain for all n . Although the increase is due to the rotational effect of advection by simple shear, the rate of increase cannot be fully explained by this rotation. Black curves in Fig. 2c and d show the angle of material lines as they are advected by unperturbed simple shear flow. It is evident that the orientations of melt bands produced in simulations and linear analysis do not track with these trajectories. The trajectories that melt bands follow have a lower rate of increase in angle with strain than do the black curves. These trajectories result from an interplay of the growth of melt bands in favoured orientation and the rotation of bands out of this orientation by simple shear. An animation of calculations with $n = 6$ (Supplementary movie M1) shows reconnection of porosity bands, maintaining the dominance of those at low angle. This reorganization is similar to the pumping mechanism proposed by Holtzman

*et al.*¹⁸, in which fluid flows from bands that have been rotated to higher angles to newly forming bands at lower angles.

Data from experiments are shown by black symbols in Fig. 2c and d and can be directly compared to model results. These data show that the peak band angle is roughly constant at $15\text{--}20^\circ$ to large strain. Up to $\gamma \approx 2$, the data are consistent with simulations with $n = 4$ or 6. Our current numerical techniques have not been able to take these simulations to higher strain (see Supplementary Information). Linear analysis, which tends to give higher peak band angles at a given strain than do simulations (see Fig. 2d), can be used to extend band-angle evolution to higher strain than that achieved in simulations (Fig. 2c). In this case, the results for $n = 6$ are most consistent with the data. However, nonlinear effects in numerical simulations such as band reconnection and realignment might lead to a lower peak band angle at a given n .

The general agreement between our calculations and experimental results helps to validate magma dynamics theory. In particular, both theory and experiment exhibit the emergence of low-angle melt bands on length scales shorter than the compaction length that persist at low angle and concentrate shear strain. This agreement demonstrates the importance of non-newtonian rheology for modelling deforming, partially molten aggregates. In detail, there are differences between experiments and theory. For example, bands produced in simulations have a smaller aspect ratio and edges that are less sharp than those in experiments. These differences may result from the specific form of the flow law that we have used. More work is needed to explore models that use other non-newtonian rheologies and other formulations of magma dynamics⁸. Nevertheless, both experiments and theory suggest that deformation-induced melt bands are a robust feature of partially molten aggregates.

These results have implications for magma transport in the mantle, for example beneath mid-ocean ridges. A fundamental observation of mid-ocean ridges is that the oceanic crust is formed within 10 km of the ridge axis, whereas melt production is believed to occur over a region extending about 100 km from the axis¹⁹. This implies some mechanism of melt focusing toward the axis. The melt-band-forming instability may contribute to melt focusing, in addition to other mechanisms that have been proposed^{20–22}. To estimate the orientation of bands, we calculate the principal axes of the strain-rate tensor for flow beneath a mid-ocean ridge²³ (Fig. 3a). Red line segments are perpendicular to the axis of maximum extension rate and show the orientation of melt bands expected for a newtonian viscosity, which tend to point away from the ridge axis. However, experiments (and theory for non-newtonian viscosity) suggest that band orientation is rotated 25° from the red line segments and should be most pronounced in regions of high strain. With these assumptions, melt bands point towards the ridge axis (Fig. 3b). If melt bands form an interconnected, permeable network with this orientation, they would enhance focusing of melt²⁴. To quantitatively test this conjecture requires large-scale magma dynamics simulations (~ 100 km) that include the effects of melt buoyancy and finite strain, with sufficiently fine spatial discretization to resolve features smaller than the compaction length ($\sim 1\text{--}10$ km)^{5,13}. Such calculations will be challenging but current results give confidence that simulations can be used to predict patterns of melt extraction from the mantle.

Received 2 January; accepted 30 June 2006.

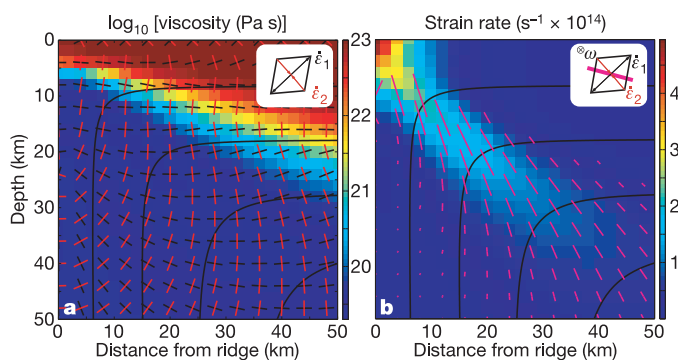


Figure 3 | Estimated orientation of melt bands beneath a mid-ocean ridge from a two-dimensional, steady-state simulation of mantle flow²³. All fields have reflection symmetry about the line $x = 0$. **a**, Viscosity (colour), flow field (black streamlines), and principal axes of the strain-rate tensor for a temperature and strain-rate-dependent rheology. Red lines are perpendicular to the direction of maximum extension. For newtonian viscosity ($n = 1$), red lines correspond to the orientation of melt bands in our model. Such bands would not focus melt toward the ridge axis. **b**, Strain rate (colour), flow field and melt-band orientation consistent with experiments (or in our model with $n \approx 4\text{--}6$). Band orientations, indicated by magenta line segments, are calculated by rotating the red lines by 25° in the direction counter to the vorticity. The length of these segments is scaled to the local strain rate. If the bands form an interconnected permeable network, this orientation could help to focus melt to the ridge axis.

1. Zimmerman, M., Zhang, S., Kohlstedt, D. & Karato, S. Melt distribution in mantle rocks deformed in shear. *Geophys. Res. Lett.* **26**, 1505–1508 (1999).
2. Holtzman, B., Groebner, N., Zimmerman, M., Ginsberg, S. & Kohlstedt, D. Stress-driven melt segregation in partially molten rocks. *Geochem. Geophys. Geosyst.* **4**, 8607 (2003).
3. Stevenson, D. Spontaneous small-scale melt segregation in partial melts undergoing deformation. *Geophys. Res. Lett.* **16**, 1067–1070 (1989).
4. Spiegelman, M. Linear analysis of melt band formation by simple shear. *Geochem. Geophys. Geosyst.* **4**, 8615, doi:10.1029/2002GC000499 (2003).
5. McKenzie, D. The generation and compaction of partially molten rock. *J. Petrol.* **25**, 713–765 (1984).

6. Scott, D. & Stevenson, D. Magma ascent by porous flow. *J. Geophys. Res.* **91**, 9283–9296 (1986).
7. Fowler, A. A mathematical model of magma transport in the asthenosphere. *Geophys. Astrophys. Fluid Dyn.* **33**, 63–96 (1985).
8. Bercovici, D., Ricard, Y. & Schubert, G. A two-phase model for compaction and damage 1. General theory. *J. Geophys. Res.* **106**, 8887–8906 (2001).
9. Hirth, G. & Kohlstedt, D. in *The Subduction Factory* (ed. Eiler, J.) 83–105 (AGU Geophysical Monograph 138, American Geophysical Union, Washington DC, 2003).
10. Richardson, C. Melt flow in a variable viscosity matrix. *Geophys. Res. Lett.* **25**, 1099–1102 (1998).
11. Hall, C. & Parmentier, E. Spontaneous melt localization in a deforming solid with viscosity variations due to water weakening. *Geophys. Res. Lett.* **27**, 9–12 (2000).
12. Karato, S. & Wu, P. Rheology of the upper mantle: A synthesis. *Science* **260**, 771–778 (1993).
13. Kelemen, P., Hirth, G., Shimizu, N., Spiegelman, M. & Dick, H. A review of melt migration processes in the adiabatically upwelling mantle beneath oceanic spreading ridges. *Phil. Trans. R. Soc. Lond. A* **355**, 283–318 (1997).
14. Mei, S., Bai, W., Hiraga, T. & Kohlstedt, D. Influence of melt on the creep behavior of olivine-basalt aggregates under hydrous conditions. *Earth Planet. Sci. Lett.* **201**, 491–507 (2002).
15. Kelemen, P., Shimizu, N. & Salters, V. Extraction of mid-ocean-ridge basalt from the upwelling mantle by focused flow of melt in dunite channels. *Nature* **375**, 747–753 (1995).
16. Knepley, M., Katz, R. & Smith, B. in *Numerical Solution of Partial Differential Equations on Parallel Computers* (eds Bruaset, A. & Tveito, A.) 413–438 (Lecture Notes in Computational Science and Engineering Vol. 51, Springer, Berlin, 2006).
17. Balay, S. *et al.* Portable Extensible Toolkit for Scientific computation (PETSc) (<http://www.mcs.anl.gov/petsc>) (2001).
18. Holtzman, B., Kohlstedt, D. & Morgan, J. P. Viscous energy dissipation and strain partitioning in partially molten rocks. *J. Petrol.* **46**, 2593–2613 (2005).
19. Forsyth, D. *et al.* Imaging the deep seismic structure beneath a mid-ocean ridge: the melt experiment. *Science* **280**, 1215–1218 (1998).
20. Spiegelman, M. & McKenzie, D. Simple 2-D models for melt extraction at mid-ocean ridges and island arcs. *Earth Planet. Sci. Lett.* **83**, 137–152 (1987).
21. Morgan, J. Melt migration beneath mid-ocean spreading centers. *Geophys. Res. Lett.* **14**, 1238–1241 (1987).
22. Sparks, D. & Parmentier, E. Melt extraction from the mantle beneath spreading centers. *Earth Planet. Sci. Lett.* **105**, 368–377 (1991).
23. Katz, R., Spiegelman, M. & Carbotte, S. Ridge migration, asthenospheric flow and the origin of magmatic segmentation in the global mid-ocean ridge system. *Geophys. Res. Lett.* **31**, L15605, doi:10.1029/2004GL020388 (2004).
24. Daines, M. & Kohlstedt, D. Influence of deformation on melt topology in peridotites. *J. Geophys. Res.* **102**, 10257–10271 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Department of Energy Computational Science Graduate Fellowship Program of the Office of Science and National Nuclear Security Administration in the Department of Energy, and the NSF. We thank B. Smith, M. Knepley and the PETSc team at Argonne National Laboratory. We are also grateful to E. Coon, P. Kelemen and the solid Earth geodynamics group at LDEO for discussions. Comments by M. Jellinek helped to improve the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.F.K. (katz@ldeo.columbia.edu).

LETTERS

Synchrotron X-ray tomographic microscopy of fossil embryos

Philip C. J. Donoghue¹, Stefan Bengtson², Xi-ping Dong³, Neil J. Gostling¹, Therese Hultgren², John A. Cunningham^{1,4}, Chongyu Yin⁵, Zhao Yue^{2,5}, Fan Peng³ & Marco Stampanoni⁶

Fossilized embryos from the late Neoproterozoic and earliest Phanerozoic have caused much excitement because they preserve the earliest stages of embryology of animals that represent the initial diversification of metazoans^{1–4}. However, the potential of this material has not been fully realized because of reliance on traditional, non-destructive methods that allow analysis of exposed surfaces only^{1–4}, and destructive methods that preserve only a single two-dimensional view of the interior of the specimen^{5,6}. Here, we have applied synchrotron-radiation X-ray tomographic microscopy (SRXTM)⁷, obtaining complete three-dimensional recordings at submicrometre resolution. The embryos are preserved by early diagenetic impregnation and encrustation with calcium phosphate, and differences in X-ray attenuation provide information about the distribution of these two diagenetic phases. Three-dimensional visualization of blastomere arrangement and diagenetic cement in cleavage embryos resolves outstanding questions about their nature, including the identity of the columnar blastomeres. The anterior and posterior anatomy of embryos of the bilaterian worm-like *Markuelia* confirms its position as a scalidophoran, providing new insights into body-plan assembly among constituent phyla. The structure of the developing germ band in another bilaterian, *Pseudoolides*, indicates a unique mode of germ-band development. SRXTM provides a method of non-invasive analysis that rivals the resolution achieved even by destructive methods, probing the very limits of fossilization and providing insight into embryology during the emergence of metazoan phyla.

The study involves various developmental stages from cleavage to pre-hatching. Fossil cleavage embryos are notoriously difficult to study because of their tiny size and precarious preservation, but

tomography data provide complete information about the size, shape, and distribution of preserved blastomeres. Thus we have been able to test the recent claim that large columnar structures in some earliest Cambrian embryos, originally interpreted as an outer

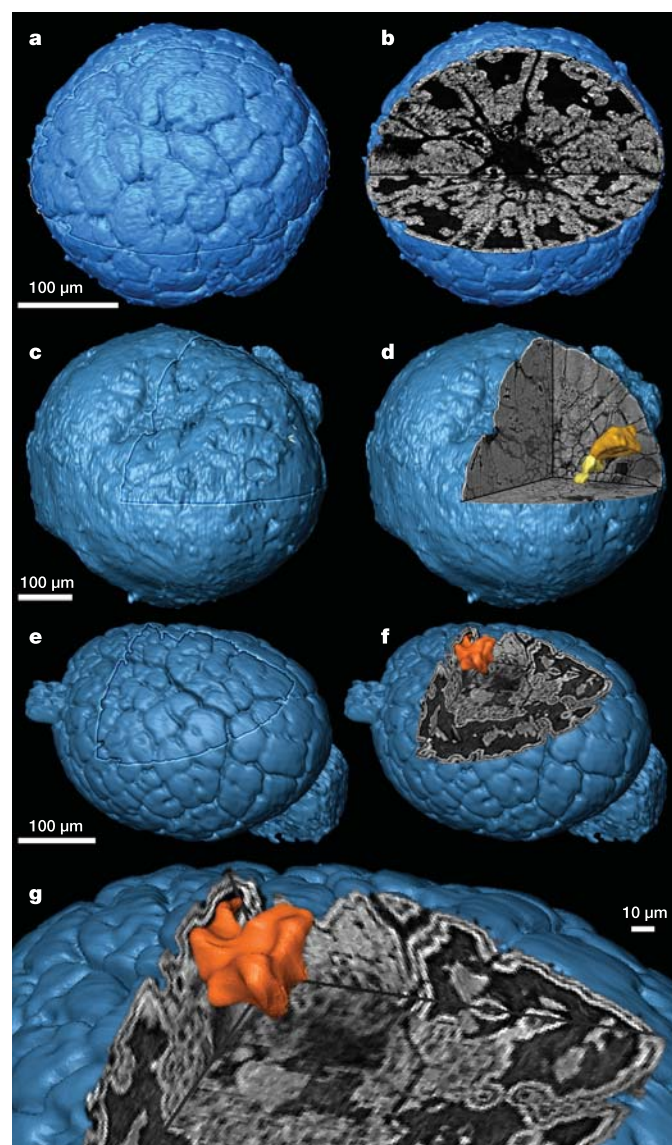


Figure 1 | Tomographic reconstructions of cleavage-stage embryos from the Lower Cambrian Kuanchuanpu Formation, Shizhonggou section, Kuanchuanpu, Ningqiang County, Shaanxi Province, Southern China. Divisions between adjacent blastomeres variably preserved on the surface and within. **a, b**, Museum of Earth Science, Institute of Geology, Chinese Academy of Geological Sciences (MESIG) 20061. Divisions between some, but not all, blastomeres are preserved internally. **c, d**, MESIG 20062. Divisions between all, or nearly all, blastomeres are preserved to their full extent; the orange and yellow structures are renderings of the morphology of a column of blastomeres. **e–g**, Geological Museum of Peking University (GMPKU) 2204. Divisions between blastomeres are generally not preserved, and instead the core of the embryo is characterized by the centrifugal addition of diagenetic crust layers (easily distinguished from edge artefacts through their absence from some of the objects seen in the slices); orange structure represents a rendering of one of the cavities within the diagenetic infilling.

¹Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK. ²Department of Palaeozoology, Swedish Museum of Natural History, Stockholm, Sweden. ³School of Earth and Space Sciences, Peking University, Beijing, China. ⁴Department of Earth and Ocean Sciences, 4 Brownlow Street, Liverpool L69 3GP, UK. ⁵Institute of Geology, Chinese Academy of Geological Sciences, Beijing, China. ⁶Swiss Light Source, Paul Scherrer Institut, CH-5232 Villigen, Switzerland.

layer of blastomeres⁸, instead represent yolk pyramids of a kind known from modern arthropods, with the implication that arthropod evolutionary history is thereby pushed back in time⁹.

Our SRXTM scans of embryos from the same Cambrian deposits show that they are fully subdivided into blastomere-like structures, among which the narrow columnar ones form the outer layer (Fig. 1a–d) (tomographic slice data available as Supplementary Information for all specimens we describe). Arthropod yolk pyramids do not have such arrangement¹⁰, and would also be expected to preserve evidence (not seen here) of the constituent yolk particles. The evidence for complete cleavage in these embryos removes one of the central arguments for associating them with the later-stage embryos of the enigmatic *Pseudoooides*³ rather than other co-occurring taxa, including the putative cnidarian *Olivoooides*^{1,8}. This also removes support for a projected prehistory of the arthropod phylum.

The SRXTM scans also illustrate the layered diagenetic crusts of apatite that often line the walls of cavities in fossilized embryos (Fig. 1g). Such layers have been interpreted^{5,6,11–14} as original cellular layers, but can now be imaged to their full three-dimensional extent (see, for example, the orange object in Fig. 1f, g), making them clearly distinguishable from cellular tissue layers. Thus, together with evidence from comparative taphonomy^{15–18}, the case is now very strong for rejecting claims of anthozoan planula larvae, hydrozoan and bilaterian gastrulae^{5,6,12}, as well as minute adult bilaterians^{12–14} from the Ediacaran Doushantuo phosphorites.

A similar mode of infilling is encountered in specimens of *Markuelia secunda*, a putative scalidophoran from the Early Cambrian of Siberia¹ (Fig. 2j). The SRXTM analysis demonstrates layered infilling, but this is clearly secondary to the replication of the

anatomy of the embryo: the two phases are distinguished by different levels of X-ray attenuation.

SRXTM has also facilitated analysis of embryonic morphological features that are not exposed at the surface. For instance, the two best-characterized species of *Markuelia* differ ostensibly in the number of pairs of posterior spines: two in *M. secunda*¹ and three in *M. hunanensis*^{4,19}. However, tomographic sections through the embryos of *M. secunda* reveal a third, unexposed pair of spines (Fig. 2j), confirming its similarity to *M. hunanensis*.

Although these data are crucial to understanding the disparity of embryo body forms, they are not germane to understanding the affinities of *Markuelia*. *Markuelia* has been proposed to fall into the Scalidophora (phyla Kinorhyncha, Loricifera, Priapulida)^{4,19}, although this conclusion is hostage to missing data, with reassignments possible within Scalidophora, or to Nematodea (phyla Nematoda, Nematomorpha)¹⁹. In particular, the organizational symmetry of the circumoral and pharyngeal spines is crucial to the systematics of these two clades, with nematoids exhibiting hexaradial symmetry and scalidophorans exhibiting 20- and 25-fold (pentaradial) symmetry²⁰. The symmetry of the spines in *Markuelia* was unknown because of the difficulty in interpreting even the very best specimens^{4,19}, but now SRXTM has made it possible to dissect such specimens virtually and resolve the scalid arrangement.

Figure 2a–f shows an incomplete embryo of *M. hunanensis* in which spines are preserved around the margins of the mouth. The exact arrangement cannot be seen in surface view, but virtual sections reveal that the spines continue internally. Their orientation towards—rather than away from—the orifice is consistent with their proposed nature as scalids situated on an introvert capable of

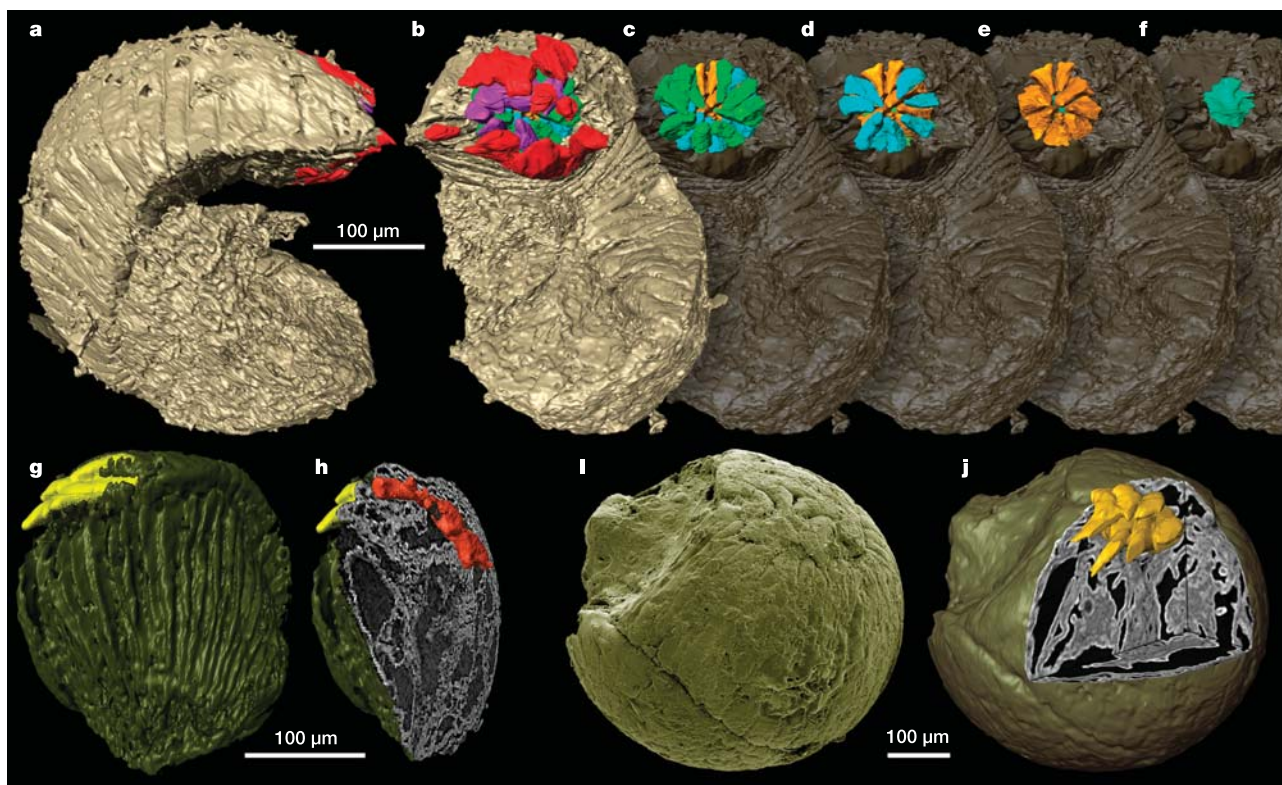


Figure 2 | The scalidophoran *Markuelia* from the Cambrian of China and Siberia. a–f, GMPKU 2205. Whole-mount tomographic reconstruction of *Markuelia hunanensis* from the Upper Cambrian of Wangcun, Hunan, Southern China. b–f are rotated 90° from a. a and b show the whole reconstruction, with all scalids. c–f, 25 circumpharyngeal scalids arranged in rings surrounding the mouth cone (f). 25 scalids are shown in the first three rings (c), 16 scalids are shown in the first two rings (d), and 8 scalids are

shown in the inner first circumoral ring (e). g, h, GMPKU2011.

Tomographic reconstruction of posterior spines and a section revealing the terminal end of the digestive tract of *M. hunanensis*. i, j, Swedish Museum of Natural History (SMNH) X2240. *Markuelia secunda* from the basal part of the Pestrotsvet Formation at Dvortsy, Aldan River, Siberia. Scanning electron micrograph (i) shows four terminal spines arranged in two pairs, while the tomographic reconstruction (j) shows a third, obscured, pair.

inversion. All but the most distal rows of scalids are located within an expansive and externally open lumen, separated from the closed pharynx by a mouth cone (Fig. 2f). Approximately forty scalids are present (including broken external scalids), arranged in five rings. The first three rings (counted away from the mouth cone) are composed of eight (Fig. 2e), eight (Fig. 2d) and nine (Fig. 2c) scalids, respectively, and are arranged such that each individual scalid intercalates with others in the preceding ring. Thus, the first three rings define 25 radii—as in living priapulids in which there are three initial rows of eight, nine and eight scalids²¹.

This overall pentameric arrangement in *Markuelia* and priapulids is also shared by loriciferans and kinorhynch, but not nematodes and nematomorphs²¹. The position of the scalids, restricted to a pre-pharyngeal lumen in the best preserved specimens, is the same in embryos and larvae of nematomorphs, kinorhynch and priapulids²², in which the introvert occurs naturally in an inverted position. The introvert normally assumes an everted position in later developmental stages and the adult^{22,23}, and it is likely that embryos of *Markuelia* in which the mouth cone is everted^{4,19,24} represent the latest-stage embryos known. We conclude, therefore, that the introvert of the *Markuelia* larva was capable of complete inversion. Together with the evidence of structure and scalid shape, it is likely that the introvert was used in peristaltic locomotion, as in the priapulid adults²³.

Data from SRXTM also reveal that the lumen at the posterior terminus, from which the three pairs of terminal spines emerge, extends internally to form an axial tube that can be traced adorally for over 200 μm (Fig. 2h). The most likely interpretation of this is a digestive tract opening to a terminal anus, a feature shared with nematodes, loriciferans, kinorhynch and basal panarthropods, but not nematomorphs or most living priapulids^{4,19}.

The inclusion of these new data on the anatomy of *Markuelia* into the phylogenetic analysis presented by Dong *et al.*^{4,19} (see Supplementary Information) adds further weight to the existing hypothesis of affinity in which *Markuelia* is resolved as a member of the scalidophoran stem, corroborating unfolding evidence that similarities to priapulids are superficial and represent shared-primitive characters of scalidophorans. Additional support is also found for the view that pentameric organization of the introvert and pharyngeal armature is a scalidophoran character²⁰, and that despite differences in adult and larval body form among scalidophorans, the primitive function of the introvert is peristaltic locomotion²⁵.

As a final example, we also examined the enigmatic fossil embryo taxon *Pseudoooides prima* (Fig. 3), which has been interpreted to exhibit a long germ-band pattern of development³. The putative germ band has been claimed to exhibit secondarily imposed metamorphism³ and is delineated from the surrounding extra-embryonic material by a marginal furrow. However, there is no evidence of progressive development, and the germ-band feature exhibits a peculiar transverse pinching at mid-length³. The SRXTM analysis reveals that in most specimens the furrows separating the presumed metameres extend internally to some depth but do not fully enclose them. The marginal furrow delineating the germ band exhibits a highly wrinkled micromorphology (Fig. 3f and Supplementary Fig. 1) but extends to only a shallow depth (Fig. 3f and the stereo Supplementary Fig. 2). Where a pinch occurs in the centre of the germ band, the marginal furrow shallows to the surface, possibly indicating the point of segment addition (Supplementary Figs 1, 2). The latter interpretation is consistent with the pattern of folding of the surface at both ends of the germ band (for example, Fig. 3a), which indicates outward movement along the band emanating from the pinched centre. A further structure, superficially like a surface scratch, is situated between the pointed ends of the germ band, at the pole opposite to its pinched centre, and oriented perpendicular to it (lower left quadrant in Fig. 3b, d; compare also Fig. 2N in ref. 3 and Fig. 4.4A in ref. 26). Tomographic slices (Fig. 3c, e) show that it is connected to internal features and is thus not merely superficial. It is

clearly of biological origin, although its nature is uncertain. We interpret it as the remains of a blastopore formed before the wrinkled surface membrane. The lower, presumably ventral, surface of the germ band is never seen intact, though it has been observed to extend to more than a radius depth within. There is no evidence to support the original interpretation of a long germ-band pattern of development³, nor of terminal segment addition, characteristic of short germ-band development. Instead, the new data suggest a unique mode of establishing metamerism in which segments were added to the germ band from a centrally located generative zone.

It is clear from this study that scanning electron microscopy and light-microscopy of thin sections are insufficient to reveal fossil embryo structure. Analyses of internal and external structure in concert by means of SRXTM have allowed us to clarify the nature of diagenetic infills, to decide between opposing interpretations of cleavage modes, and to resolve the anatomy of the later-stage embryos of *Pseudoooides* and *Markuelia*, helping to constrain their affinity and evolutionary significance. Perhaps more importantly, it

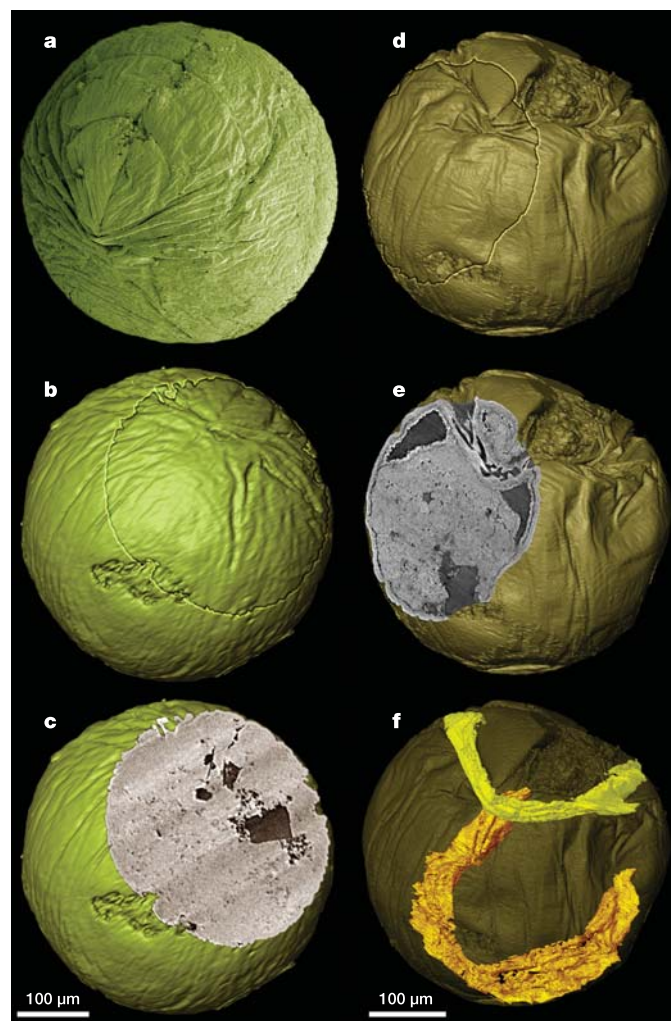


Figure 3 | Germ-band embryos of *Pseudoooides prima*, same locality as for Fig. 1. a–c, MESIG 20063. a, Scanning electron micrograph. b, c, Tomographic reconstruction with virtual histological section (c) showing diagenetic cuboidal crystals and possible blastopore (below and left of centre) at opposite pole from the germ band. d–f, MESIG 20065. Tomographic reconstruction, oriented with germ band uppermost. d shows position of virtual section in e, which shows that the terminal tip has an extension of the marginal furrow below it and that the possible blastopore is connected to a lacuna. f, rendered cast of the marginal furrow revealed as broadly horseshoe-shaped, enclosing the pointed ends of the germ band.

has revealed aspects of the anatomy of these organisms that would never have otherwise been resolved.

While authigenic phosphatisation of microscopic organisms has been widely recognized as an exceptional mode of fossilization²⁷, it has been held that preservation is limited to surface structure²⁸. SRXTM confirms that at least in instances of embryo fossilization, preservation can be much more extensive. More likely, the extent of anatomical preservation in fossils such as those from the Orsten and Orsten-like deposits²⁸ has been significantly underestimated and warrants further investigation with SRXTM technology.

SRXTM provides a non-invasive method of analysis of small and microscopic fossil materials that rivals even destructive approaches in terms of the resolution of data recovered, unlocking the finest details of preserved anatomy from fossilized remains. The present study demonstrates the feasibility of the method for a variety of questions concerning developmental processes in early fossil animals. The method has wide applicability in the study of microscopic structures in originally mineralized and exceptionally preserved fossil materials and may thus bring about a revolution in palaeontology on a par with that once brought about by the scanning electron microscope.

Received 21 February; accepted 10 May 2006.

- Bengtson, S. & Yue, Z. Fossilized metazoan embryos from the earliest Cambrian. *Science* **277**, 1645–1648 (1997).
- Xiao, S., Zhang, Y. & Knoll, A. H. Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphate. *Nature* **391**, 553–558 (1998).
- Steiner, M., Zhu, M., Li, G., Qian, Y. & Erdtmann, B.-D. New early Cambrian bilaterian embryos and larvae from China. *Geology* **32**, 833–836 (2004).
- Dong, X., Donoghue, P. C. J., Cheng, H. & Liu, J. Fossil embryos from the Middle and Late Cambrian period of Hunan, south China. *Nature* **427**, 237–240 (2004).
- Chen, J. *et al.* Precambrian animal diversity: putative phosphatised embryos from the Doushantuo Formation of China. *Proc. Natl Acad. Sci. USA* **97**, 4457–4462 (2000).
- Chen, J. *et al.* Precambrian animal life: Probable developmental and adult cnidarian forms from Southwest China. *Dev. Biol.* **248**, 182–196 (2002).
- Stampanoni, M. *et al.* High resolution X-ray detector for synchrotron-based microtomography. *Nucl. Instrum. Methods Phys. Res. A* **491**, 291–301 (2002).
- Yue, Z. & Bengtson, S. Embryonic and post-embryonic development of the Early Cambrian cnidarian *Olivoides*. *Lethaia* **32**, 181–195 (1999).
- Chen, J., Braun, A., Waloszek, D., Peng, Q.-Q. & Maas, A. Lower Cambrian yolk-pyramid embryos from southern Shaanxi, China. *Prog. Nat. Sci.* **14**, 167–172 (2004).
- Chipman, A. D., Arthur, W. & Akam, M. Early development and segment formation in the centipede, *Strigamia maritima* (Geophilomorpha). *Evol. Dev.* **6**, 78–89 (2004).
- Chen, J. Y. & Chi, H. M. Precambrian phosphatized embryos and larvae from the Doushantuo Formation and their affinities, Guizhou (SW China). *Chin. Sci. Bull.* **50**, 2193–2200 (2005).
- Chen, J. Y. *The Dawn Of Animal World* 1–366 (Jiangsu Science and Technology Publishing House, Nanjing, 2004).
- Chen, J. *et al.* Small bilaterian fossils from 40 to 55 million years before the Cambrian. *Science* **305**, 218–222 (2004).
- Chen, J.-Y., Oliveri, P., Davidson, E. & Bottjer, D. J. Response to ‘Comment on “Small bilaterian fossils from 40 to 55 million years before the Cambrian”’. *Science* **306**, 1291b (2004).
- Bengtson, S. & Budd, G. Comment on “Small bilaterian fossils from 40 to 55 million years before the Cambrian”. *Science* **306**, 1291a (2004).
- Bengtson, S. in *The New Panorama Of Animal Evolution* (eds Legakis, A., Sfenthourakis, S., Polymeni, R. & Thessalou-Legaki, M.) 289–300 (Pensoft, Moscow, 2003).
- Xiao, S., Yuan, X. & Knoll, A. H. Eumetazoan fossils in terminal Proterozoic phosphorites? *Proc. Natl Acad. Sci. USA* **97**, 13684–13689 (2000).
- Raff, E. C., Vilinski, J. T., Turner, F. R., Donoghue, P. C. J. & Raff, R. A. Experimental taphonomy shows the feasibility of fossil embryos. *Proc. Natl Acad. Sci. USA* **103**, 5846–5851 (2006).
- Dong, X., Donoghue, P. C. J., Cunningham, J., Liu, J. & Cheng, H. The anatomy, affinity and phylogenetic significance of Markuelia. *Evol. Dev.* **7**, 468–482 (2005).
- Schmidt-Rhaesa, A. Phylogenetic relationships of the Nematomorpha—a discussion of current hypotheses. *Zool. Anz.* **236**, 203–216 (1998).
- Lemburg, C. *Ultrastrukturelle Untersuchungen an den Larven von Halicyptus spinulosus und Priapulus caudatus: Hypothesen zur Phylogenie der Priapulida und deren Bedeutung für die Evolution der Nemathelminthes* 1–393 (Cuvillier, Göttingen, 1999).
- Adrianov, A. V. & Malakhov, V. V. *Kinorhyncha: Structure, Development, Phylogeny and Taxonomy* 1–260 (Nauka Publishing, Moscow, 1994).
- Hammond, R. A. The burrowing of *Priapulus caudatus*. *J. Zool.* **162**, 469–480 (1970).
- Donoghue, P. C. J. *et al.* Fossilized embryos are widespread but the record is temporally and taxonomically biased. *Evol. Dev.* **8**, 232–238 (2006).
- Dzik, J. & Krumbiegel, G. The oldest ‘onychophoran’ *Xenusion*: a link connecting phyla? *Lethaia* **22**, 169–182 (1989).
- Steiner, M., Li, G. & Zhu, M. Lower Cambrian small shelly fossils of northern Sichuan and southern Shaanxi (China), and their biostratigraphic significance. *Geobios* **37**, 259–275 (2004).
- Briggs, D. E. G. The role of decay and mineralization in the preservation of soft-bodied fossils. *Annu. Rev. Earth Planet. Sci.* **31**, 275–301 (2003).
- Waloszek, D. The ‘Orsten’ window—a three-dimensionally preserved Upper Cambrian meiofauna and its contribution to our understanding of the evolution of Arthropoda. *Paleontol. Res.* **7**, 71–88 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. Guo for field assistance, and A. Groso for her support at the beamline, and K. Robson-Brown for her assistance in running the experiments. This work was funded by the Swiss Light Source (P.C.J.D. and S.B.), European Union FP6, as well as grants from the Natural Environment Research Council (P.C.J.D.), the Swedish Research Council (S.B.), Bristol University and the National Natural Science Foundation of China (X.-p.D.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.C.J.D. (phil.donoghue@bristol.ac.uk) or M.S. (marco.stampanoni@psi.ch).

LETTERS

Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements

D. Chourrout¹, F. Delsuc², P. Chourrout³, R. B. Edvardsen¹, F. Rentzsch¹, E. Renfer¹, M. F. Jensen¹, B. Zhu⁴, P. de Jong⁴, R. E. Steele⁵ & U. Technau¹

Bilaterian animals have a Hox gene cluster essential for patterning the main body axis, and a ParaHox gene cluster. Comparison of Hox and ParaHox genes has led workers to postulate that both clusters originated from the duplication of an ancient cluster named ProtoHox, which contained up to four genes with at least the precursors of anterior and posterior Hox/ParaHox genes^{1–3}. However, the way in which genes diversified within the ProtoHox, Hox and ParaHox clusters remains unclear because no systematic study of non-bilaterian animals exists. Here we characterize the full Hox/ParaHox gene complements and genomic organization in two cnidarian species (*Nematostella vectensis* and *Hydra magnipapillata*), and suggest a ProtoHox cluster simpler than originally thought on the basis of three arguments. First, both species possess bilaterian-like anterior Hox genes, but their non-anterior genes do not appear as counterparts of either bilaterian central or posterior genes; second, two clustered ParaHox genes, *Gsx* and a gene related to *Xlox* and *Cdx*, are found in *Nematostella vectensis*; and third, we do not find clear phylogenetic support for a common origin of bilaterian *Cdx* and posterior genes, which might therefore have appeared after the ProtoHox cluster

duplication. Consequently, the ProtoHox cluster might have consisted of only two anterior genes. Non-anterior genes could have appeared independently in the Hox and ParaHox clusters, possibly after the separation of bilaterians and cnidarians.

Recent molecular phylogenies support the contention that the non-bilaterian Cnidaria is the sister group to bilaterians^{4,5} and can therefore provide information for reconstructing the early history of bilaterian homeobox gene complements. Earlier searches for cnidarian homeobox genes have revealed the presence of anterior-like Hox genes beside *Gsx*, so that the ProtoHox cluster must have been duplicated before the cnidarian–bilaterian split. Previous reports^{6–13} have proposed that posterior Hox genes, but not central genes or *Hox3*, are also present in cnidarians. We used the publicly available high-coverage genome shotgun sequences to identify the complete set of homeobox genes of two distantly related cnidarians, the freshwater polyp *Hydra magnipapillata* (Hydrozoa) and the sea anemone *Nematostella vectensis* (Anthozoa). *Nematostella* is particularly informative because it is considered to represent the basal group within the Cnidaria^{14,15}.

Eighteen *Nematostella* candidate homeodomains were allocated to

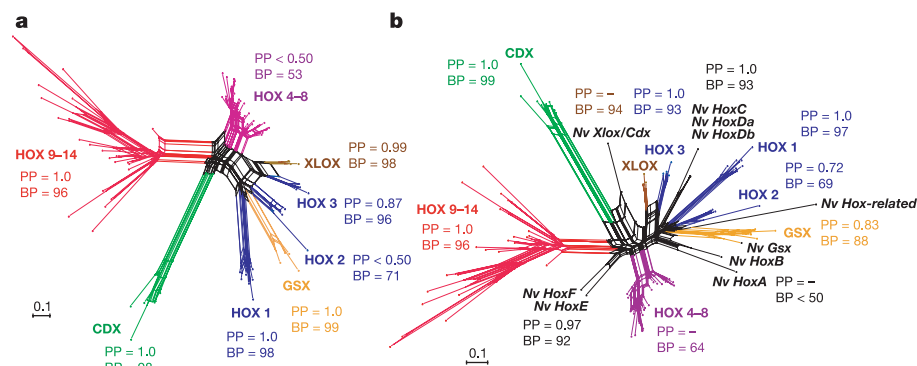


Figure 1 | Phylogenetic analyses of bilaterian and *Nematostella* Hox genes based on their homeodomain. **a**, Neighbour-net network of 82 bilaterian Hox and ParaHox genes based on maximum-likelihood (ML) distances estimated with a Jones–Taylor–Thornton (JTT) substitution matrix plus an eight-category gamma rate heterogeneity correction (JTT + Γ_8). Note the large differences in evolutionary rates between the fast-evolving posterior Hox and Cdx genes and the slowly evolving central Hox genes. Bayesian posterior probabilities (PP) for the major groups have been mapped on the network. A dash (–) indicates that the corresponding split was not observed in the Bayesian analysis. Bootstrap proportions (BP) were inferred from the BioNJ algorithm for the same data set and JTT + Γ_8 ML distance estimates.

The scale bar represents the estimated number of changes per site. Note the position of the Cdx group, which is distinct from the posterior Hox group. **b**, Neighbour-net network of 92 Hox and ParaHox genes from bilaterian and *Nematostella* based on JTT + Γ_8 ML distance estimates. Note the position of non-anterior *Nematostella* genes (*Nv HoxE* and *Nv HoxF*). The major groups of bilaterian Hox and ParaHox proteins remain unaffected by the addition of the *Nematostella* data set. Bayesian posterior probabilities (PP) and bootstrap proportions (BP) were calculated as for the bilaterian data set. More detailed figures, including all sequence origins, are available in Supplementary Information.

¹Sars International Centre for Marine Molecular Biology, University of Bergen, Thormøhlensgt. 55, 5008 Bergen, Norway. ²Laboratoire de Paléontologie, Phylogénie et Paléobiologie, Institut des Sciences de l'Évolution, UMR5554-CNRS, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France. ³American Hospital of Paris, 63, boulevard Victor Hugo, BP 109, 92202 Neuilly-sur-Seine Cedex, France. ⁴Children's Hospital and Research Center at Oakland, Oakland, California 94609, USA. ⁵Department of Biological Chemistry, University of California, Irvine, 240D Medical Sciences I, Irvine, California 92697-1700, USA.

the Antennapedia class/Hox subclass and named on the basis of their relationship with vertebrate and fly sequences. Nine of them were clearly not Hox genes; these included the ParaHox gene *Gsx*, four distinct *Mox* genes, and *Mnx*, *Gbx*, *Rough* and *Evx*. Seven were considered as candidate Hox genes, including five genes already known¹³, which we renamed to avoid confusions with bilaterian Hox paralogues: *HoxA* (formerly *anthox6*), *HoxB* (previously unknown), *HoxC* (formerly *anthox7*), *HoxDa* and *HoxDb* (two distinct copies with an identical homeobox of a gene known as *anthox8*), *HoxE* (formerly *anthox1a*) and *HoxF* (formerly *anthox1*). Finally, two previously unknown genes showed higher divergence from Antp/Hox genes of other species (see below). In contrast, a search in the *Hydra* genome revealed only eight Antp/Hox genes, including one *Mox* gene, one *Gsx* gene and six candidate Hox genes.

To clarify the relationships between bilaterian and cnidarian Hox/ParaHox genes, we performed phylogenetic analyses (Supplementary Methods), including the evaluation of alternative hypotheses that receive some support in the data but would not be revealed with traditional tree-building methods. Most reliable trees were obtained with large data sets of homeodomain proteins from putative slowly evolving bilaterian species such as *Nereis virens*,

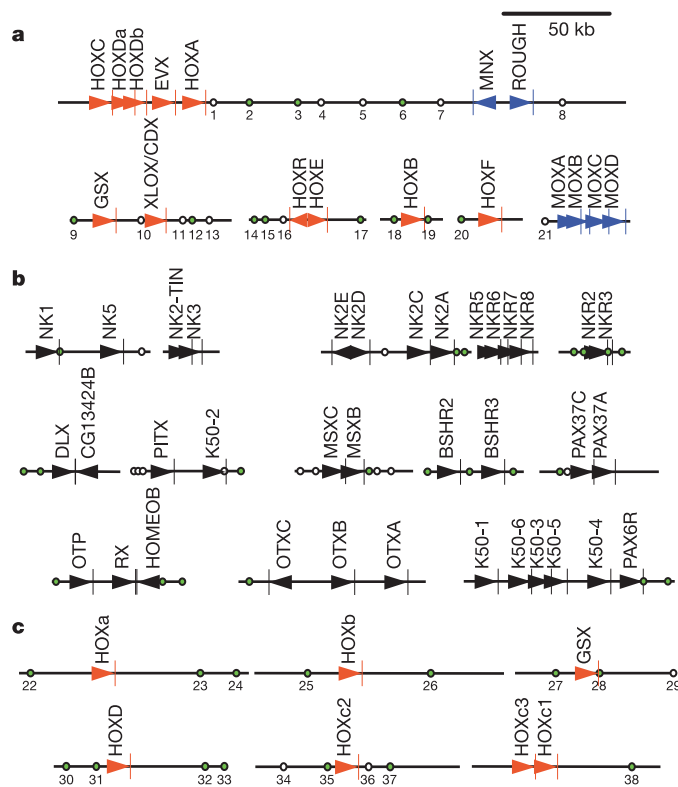


Figure 2 | All physical linkages detected between cnidarian homeobox genes. **a**, Linkages associating *Nematostella* ANTP/Hox-like genes were validated by BAC library screening with individual Hox-like genes, before an *in silico* genome walk using the local assembler of shotgun traces allowed us to reconstruct the cluster DNA sequences. The positions of Hox and Hox-related genes including the ParaHox gene *Gsx* and their respective orientations are represented by red arrows, and other linked homeobox genes by blue arrows. Putative intervening non-homeobox genes are represented by circles, filled with green when expression is confirmed by EST alignments; sequence similarity with known genes was detected by BLASTX (Supplementary Information). **b**, Linkages associating other *Nematostella* homeobox genes were validated with an *in silico* genome walk using the local assembler of shotgun traces. **c**, Linkages associating *Hydra* ANTP/Hox-like genes including *Gsx* were validated with an *in silico* genome walk. Genes found in their environments have been annotated by using BLASTX and alignments with *Hydra* ESTs (Supplementary Information).

Cupiennius salei, *Ptychodera flava* and *Branchiostoma floridae*. First, with the use of bilaterian sequences only, we were able to recognize the major groups of Hox and ParaHox genes (Fig. 1a and Supplementary Information) and confirm that ParaHox genes *Gsx* and *Xlox* cluster with the anterior genes *Hox1/2* and *Hox3*, respectively. Surprisingly, the group of ParaHox *Cdx* genes seemed divergent from all three groups of bilaterian Hox genes and showed only slight phylogenetic affinity (bootstrap proportions = 38) with posterior Hox genes (Fig. 1a and Supplementary Information). The existence of a precursor for Hox posterior genes and *Cdx* in the ProtoHox cluster is therefore questionable.

Neighbour-net analyses clearly showed that within cnidarians *Nematostella* is the slowest evolving species, and the *Nematostella* data set was consequently used for extensive sequence comparisons. *HoxA*, *HoxB*, *HoxC*, *HoxDa/Db* and one divergent gene (*Hox-related*) clustered with the anterior genes (*Hox1–3*) of bilaterians (Fig. 1b and Supplementary Information). The precise relationships within the anterior Hox homeodomains could not be statistically resolved despite weak preferential affinities of *HoxA/HoxB* with *Hox1/2*, and of *HoxC/HoxDa/HoxDb* with *Hox3* in several analyses (not shown). By comparison, *HoxE* and *HoxF*, which were previously tentatively classified as posterior genes¹³, did not preferentially group with anterior, central or posterior genes (Fig. 1b and Supplementary Information). Instead, they formed an independent and strongly supported branch. Last, one of the two divergent genes showed conflicting affinities in network analyses with *Cdx* and *Xlox* (Fig. 1b) and is therefore most probably a ParaHox gene. This gene, named *Xlox/Cdx*, is not a second Gsx gene, according to likelihood-based

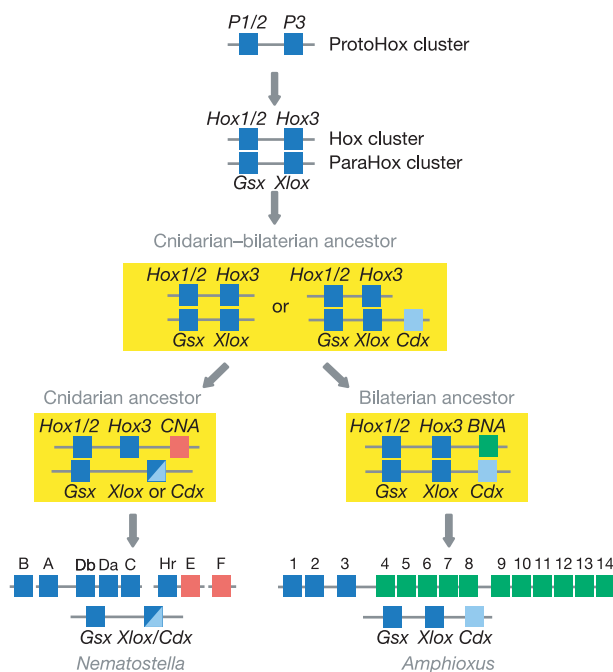


Figure 3 | Parsimonious consensus model for the evolution of Hox/ParaHox clusters of bilaterians. The ProtoHox cluster may have contained only two anterior genes (*P1/2* and *P3*) and its duplication generated two equally simple Hox and ParaHox clusters. The prototypal bilaterian Hox cluster may have contained an extra non-anterior Hox gene (*BNA*), a precursor of future central and posterior genes. An independent duplication in the Hox cluster would have generated the precursor of cnidarian non-anterior Hox genes (*CNA*), which later on became HoxE and HoxF of *Nematostella*. The third ParaHox gene *Cdx* might have appeared through a duplication in the ParaHox cluster, and not in the ProtoHox cluster as generally proposed, either in early bilaterians or before the split of bilaterians and cnidarians. In the latter case, cnidarians had already simplified their ParaHox cluster through the loss of *Xlox*.

statistical tests (Supplementary Information). When Hox genes from other cnidarians were included, the same picture appeared with cnidarian-specific non-anterior Hox genes, although their long branches tended to be attracted towards the fast-evolving posterior genes (Supplementary Information). Overall, cnidarians possess genuine anterior Hox genes, several non-anterior Hox genes not clearly related to central and posterior genes of bilaterians, and *Gsx* and a second candidate ParaHox gene related to *Xlox* and *Cdx*.

Using bacterial artificial chromosome (BAC) library screening and sequence reconstruction from shotgun data, we established linkages between distinct homeobox genes in the *Nematostella* genome. The results (Fig. 2a and Supplementary Information) showed a main cluster of five genes (*HoxC/HoxDa/HoxDb/Evx/HoxA*) in a tandem array over about 50 kilobases (kb). Linkage between the non-Hox gene *Evx* and a Hox gene has previously been shown in another anthozoan cnidarian¹⁶. We extended the cluster sequence and reached two other homeobox genes about 200 kb downstream: one is *Mnx*, a gene also present in the neighbourhood of chordate Hox clusters¹⁷, and the other is the orthologue of the fly gene *Rough*, which is on the same chromosome arm, 3R, as the ANT-C and BX-C complexes (<http://www.ncbi.nlm.nih.gov/projects/genome/guide/fly/>). Another linkage was established between the anterior gene *Hox-related* and the non-anterior gene *HoxE*, whereas *HoxB* and *HoxF* seemed to be isolated in two distinct genome locations. In sum, linkages in the *Nematostella* genome provide support for the existence of an ancient cnidarian Hox cluster associating *Evx*, several anterior Hox genes and at least one non-anterior Hox gene. In the *Nematostella* lineage, the cluster was expanded through gene duplications. Our analyses robustly support the proposition that two precursor genes have been amplified into *HoxC/HoxDa/HoxDb* and into *HoxE/HoxF* subsets, respectively. A common origin for *HoxA* and *HoxB* is also indicated by the analyses, although with weaker support. The cluster was also affected by a rearrangement placing *Evx* between anterior Hox genes, and by at least two splits. Our linkage studies placed the candidate *Xlox/Cdx* gene immediately downstream of *Gsx*, in tandem orientation. Thus, this association can represent a ParaHox cluster. By comparison, genomic sequence reconstruction of the *Hydra* Hox gene complement provided evidence for strong degeneration of the Hox cluster in *Hydra*: only two non-anterior genes, most probably resulting from a recent duplication (*HmHoxc1* and *HmHoxc3*), could be linked (Fig. 2c), and *Gsx* was the only ParaHox gene detected.

As lineage-specific duplications in the Hox cluster are rare within bilaterians, we wanted to know whether or not those found in *Nematostella* are limited to the Hox genes. For this we examined potential linkages between all other homeobox genes. Including the ten Hox-like genes mentioned above, we found a total of 139 homeobox genes, a surprisingly high number in comparison with other invertebrates (Table 1 and Supplementary Information). Phylogenetic analyses allowed the placing of at least 87 of those into 58 known groups of homeobox genes, out of 76 groups known for bilaterians. They also indicated that 42 unclassified genes might have arisen through recent amplifications of at most ten genes. By comparison, *Hydra* has a considerably smaller number of homeobox genes and gene groups. Because all *Hydra* homeobox gene groups have representatives in the *Nematostella* genome, we assume that the homeobox complement has been greatly reduced in the *Hydra* lineage (Table 1). Comparisons between the extended homeobox sequences of *Nematostella* allowed the detection of an additional 13 physical clusters (Fig. 2b and Supplementary Information). Four of these clusters may have a more ancient origin, either because they have been identified in the genomes of bilaterians or because they associate distantly related genes. The other nine clusters are undoubtedly the result of recent tandem duplications. We also did not detect obvious synteny conservation between the environments of well related but unlinked homeobox genes, as might be expected after whole-genome duplication (not shown). Hence the gene duplications observed in the *Nematostella* Hox genes represent a general phenomenon for the homeobox gene complement.

From our data, the most parsimonious model for the evolution of Hox/ParaHox clusters is as follows (Fig. 3). Two ProtoHox genes (*P1/2* and *P3*) gave rise to the Hox cluster consisting of two anterior Hox genes *Hox1/2* and *Hox3*, and to the ParaHox cluster with *Gsx* and *Xlox*. Subsequently, internal duplications expanded the Hox cluster, first by adding a precursor of the non-anterior genes. This gene might have appeared before the cnidarian–bilaterian split or independently in both lineages. Similarly, *Cdx* was added in the ParaHox cluster, either in the bilaterian lineage or before the cnidarian–bilaterian split. After this split, the cnidarian and bilaterian Hox clusters were further expanded through lineage-specific duplications. In this model, cnidarians never had genuine central and posterior Hox genes, and perhaps no *Cdx* gene either. Alternative models can be proposed but they seem to be less parsimonious. For example, the

Table 1 | Classification of homeobox genes from *Nematostella vectensis* and *Hydra magnipapillata* based on phylogenetic analyses

Class/subclass	Superfamily	No. of gene groups (without Hox/ParaHox)			No. of genes (with Hox/ParaHox)	
		Known	<i>Nematostella</i>	<i>Hydra</i>	<i>Nematostella</i>	<i>Hydra</i>
Antp/Hox-like	Hox	–	–	–	8	6
	ParaHox	–	–	–	2	1
	Extended-Hox	2	2	1	5	1
	EHG-box	3	3		3	
Antp/NK-like	Placed	20	18	8	35	11
	Unplaced				25	
Prd/Paired-like Q50	Placed	15	9	3	9	5
	Unplaced				8	3
Prd/Paired-like K50	Placed	4	3	3	5	5
	Unplaced				6	1
Prd/Pax S50		4	3	1–2	5	3
Pou		6	4	2	5	2
Lim		6	6	3–4	6	4
Tale		5	5	3	7	4–5
Six		3	3	2	5	2
Prox		1				
Cut		3	1		1	
Zfh		2				
Hnf1		2	1		1	
Unplaced					3	4
Total		76	58	26–28	139	52–53

In *Nematostella*, a significant subset of genes could be allocated to known classes and superfamilies but diverged from known homeobox gene groups, while three genes could not be allocated to known classes (see also Supplementary Information). The *Hydra* genome shows a substantial reduction of the homeobox gene complement, and all identified gene groups are also present in *Nematostella*.

precursor of non-anterior genes could have already existed in the ProtoHox cluster but had then to evolve in multiple directions, leading to central/posterior Hox genes of bilaterians, *HoxE/F* of cnidarians, and *Cdx*. Alternatively, it is possible that there were even more such precursors in the ProtoHox cluster and some were lost in one lineage and/or in one of the two clusters.

Another main angle from which to interpret the existing data is that cnidarians are in fact simplified bilaterians, with a reduced mesoderm as one of the possible regressions^{13,18}. A loss of central and posterior genes, or their rapid divergence into current non-anterior Hox genes of cnidarians, could have accompanied the process of simplification, as the Hox cluster breakdowns also would have done. Although this cannot be ruled out completely, extensive expressed sequence tag (EST)¹⁹ and genome surveys in *Nematostella* identified a surprising complexity of genes, including many genes lost in bilaterian lineages. This and the considerable degree of gene sequence conservation of the *Nematostella* lineage argues against a highly reduced and derived organism. In addition, more complex Hox gene complements in more basal metazoan phyla such as Placozoa, which would be indicative of gene loss in Cnidaria, could not be found²⁰. We therefore argue that the common ancestor of cnidarians and bilaterians contained rather simple primordial Hox and ParaHox clusters that had distinct fates in Cnidaria and Bilateria. The ProtoHox cluster itself might have consisted of only two anterior genes.

METHODS

Homeodomains were searched within each set of cnidarian genome data with the use of TBLASTN at low stringency (10^{-8}) with representative query sequences from mouse and the other cnidarian and classified into known classes, subclasses and groups. Alignments of shotgun data sets with ESTs had indicated that the probability of a given gene's being represented exceeded 98%. Phylogenetic analyses were based on maximum-likelihood, bayesian and distance-based network analyses of homeodomain proteins. Phylogenetic network analyses allowed us to detect conflicting signals and areas of uncertainty in the data, which appear with homeodomains as a result of reduced alignment lengths. Genomic sequences were extended by using the in-house-developed software MâD (Marche à Droite), which performs walks and jumps using shotgun sequences and their physical links. Linkages between homeobox genes were established through BAC library screening and confirmed with polymerase chain reaction on positive BAC clones, with MâD reconstruction of long sequence contigs. All technical details are provided in Supplementary Information.

Received 11 January; accepted 8 May 2006.

1. Ferrier, D. E. K. & Holland, P. W. Ancient origin of the Hox gene cluster. *Nature Rev. Genet.* **2**, 33–38 (2001).
2. Garcia-Fernandez, J. The genesis and evolution of homeobox gene clusters. *Nature Rev. Genet.* **6**, 881–892 (2005).
3. Brooke, N. M., Garcia-Fernandez, J. & Holland, P. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* **392**, 920–922 (1998).
4. Collins, A. G. Evaluating multiple alternative hypotheses for the origin of Bilateria: an analysis of 18S molecular evidence. *Proc. Natl Acad. Sci. USA* **95**, 15458–15463 (1998).
5. Philippe, H., Lartillot, N. & Brinkmann, H. Multigene analysis of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa and Protostomia. *Mol. Biol. Evol.* **22**, 1246–1253 (2005).
6. Shenk, M. A., Bode, H. R. & Steele, R. E. Expression of *Cnox-2*, a HOM/HOX homeobox gene in hydra, is correlated with axial pattern formation. *Development* **117**, 657–667 (1993).
7. Gauchat, D. *et al.* Evolution of Antp-class genes and differential expression of Hydra Hox/paraHox genes in anterior patterning. *Proc. Natl Acad. Sci. USA* **97**, 4493–4498 (2000).
8. Finnerty, J. R. & Martindale, M. Q. Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria. *Evol. Dev.* **1**, 16–23 (1999).
9. Finnerty, J. R. *et al.* Early evolution of a homeobox gene: the parahox gene *Gsx* in the Cnidaria and the Bilateria. *Evol. Dev.* **5**, 331–345 (2003).
10. Kuhn, K., Streit, B. & Schierwater, B. Homeobox genes in the cnidarian *Eleuthera dichotoma*: evolutionary implications for the origin of Antennapedia-class (HOM/Hox) genes. *Mol. Phylogenet. Evol.* **6**, 30–38 (1996).
11. Kuhn, K., Streit, B. & Schierwater, B. Isolation of Hox genes from the scyphozoan *Cassiopeia xamachana*: implications for the early evolution of Hox genes. *J. Exp. Zool.* **285**, 63–75 (1999).
12. Martinez, D. A. *et al.* Cnidarian homeoboxes and the zootype. *Nature* **393**, 748–749 (1998).
13. Finnerty, J. R. *et al.* Origins of bilateral symmetry: *Hox* and *Dpp* expression in a sea anemone. *Science* **304**, 1335–1337 (2004).
14. Collins, A. G. Phylogeny of the Medusozoa and the evolution of the cnidarian life cycles. *J. Evol. Biol.* **15**, 418–432 (2002).
15. Darling, J. A. *et al.* Rising starlet: the starlet sea anemone, *Nematostella vectensis*. *Bioessays* **27**, 211–221 (2005).
16. Miller, D. J. & Miles, A. Homeobox genes and the zootype. *Nature* **365**, 215–216 (1993).
17. Castro, L. F. C. & Holland, P. W. H. Chromosomal mapping of ANTP class homeobox genes in *Amphioxus*: piecing together ancestral genomes. *Evol. Dev.* **5**, 459–465 (2003).
18. Seipel, K. & Schmid, V. Evolution of striated muscle: jellyfish and the origin of triploblasty. *Dev. Biol.* **282**, 14–26 (2005).
19. Technau, U. *et al.* Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends Genet.* **21**, 633–639 (2005).
20. Monteiro, A. S., Schierwater, B., Dellaporta, S. L. & Holland, P. W. H. A low diversity of ANTP class homeobox genes in Placozoa. *Evol. Dev.* **8**, 174–182 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Rokhsar for his efforts in heading the *Nematostella* Genome project at DoE-JGI, and the J. Craig Venter Institute and the National Human Genome Research Institute for the *Hydra* Genome Project. The initial EST screen in *Nematostella* was funded by the Deutsche Forschungsgemeinschaft, and the construction of the *Nematostella* BAC library was funded by the National Science Foundation (NSF). This is contribution ISEM-2006-37 of the Institut des Sciences de l'Evolution de Montpellier.

Author Contributions The project was conceived and the manuscript written by D.C. and U.T. The bioinformatic identification and genomic assemblies of the homeobox gene complement in the *Hydra* and *Nematostella* genome datasets were conducted by D.C. The software package including MâD was written by P.C., who conceived and trained it with D.C. The phylogenetic analyses were performed by F.D. R.B.E. also participated in the homeobox gene classification and phylogenetic analysis. The cloning of cDNAs and genomic clones of Hox-related genes was performed by U.T., E.R. and M.F.J.; BAC library screening was conducted by U.T. and E.R.; and BAC colony PCR was performed by M.F.J. The *in situ* hybridizations were performed by F.R. R.E.S. and U.T. are members of the *Nematostella* genome consortium, and R.E.S. is a member of the *Hydra* genome consortium. R.E.S., P.deJ. and U.T. were co-principal investigators on the NSF grant for the generation of the *Nematostella* BAC library; U.T. collected the animal material and prepared the DNA; and B.Z. generated the BAC library under the supervision of P.deJ.

Author Information Sequences have been deposited in GenBank with accession numbers DQ500742–DQ500879. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.C. (Daniel.chourrout@sars.uib.no) and U.T. (Ulrich.technau@sars.uib.no).

LETTERS

Mesodermal Wnt2b signalling positively regulates liver specification

Elke A. Ober^{1†}, Heather Verkade^{1†}, Holly A. Field^{1†} & Didier Y. R. Stainier¹

Endodermal organs such as the lung, liver and pancreas emerge at precise locations along the primitive gut tube. Although several signalling pathways have been implicated in liver formation^{1,2}, so far no single gene has been identified that exclusively regulates liver specification. In zebrafish, the onset of liver specification is marked by the localized endodermal expression of *hhex* and *prox1* at 22 hours post fertilization. Here we used a screen for mutations affecting endodermal organ morphogenesis to identify a unique phenotype: *prometheus* (*prt*) mutants exhibit profound, though transient, defects in liver specification. Positional cloning reveals that *prt* encodes a previously unidentified Wnt2b homologue. *prt/wnt2bb* is expressed in restricted bilateral domains in the lateral plate mesoderm directly adjacent to the liver-forming endoderm. Mosaic analyses show the requirement for *Prt/Wnt2bb* in the lateral plate mesoderm, in agreement with the inductive properties of Wnt signalling. Taken together, these data reveal an unexpected positive role for Wnt signalling in liver specification, and indicate a possible common theme for the localized formation of endodermal organs along the gut tube.

During development of the digestive system, the different accessory organs such as the lung, the liver and the dorsal and ventral pancreatic anlagen form at specific positions along the anteroposterior axis of the digestive tract. The precursors for these endodermal organs are located in close proximity to each other, requiring a highly regulated network of signals to specify each population. Although the molecular details of liver differentiation are beginning to emerge^{1–4}, the processes leading to its local induction are still poorly understood. This gap in our knowledge is reflected by the absence of mutations that specifically interfere with the initial specification of the liver anlage.

To identify factors involved in endodermal organ morphogenesis, we performed a forward genetic screen (E.A.O., H.V., H.A.F., P. D. Si Dong, P. Aanstad, T. Sakaguchi, M. Bagnat, C. Munson, W.-S. Chung, C. Shin, S. Curado, R. Anderson, J. Frantsve, D. Beis, T. Bartman and D.Y.R.S., unpublished observations) using the *Tg(gutGFP)*^{s854} line, a transgenic zebrafish line expressing green fluorescent protein (GFP) throughout the developing endoderm⁵. We identified the recessive mutant *prt* in which, at 50 hours post fertilization (h.p.f.), the liver is absent (Fig. 1b) or strongly reduced (Fig. 1c) compared to wild-type siblings of the same stage (Fig. 1a). Three phenotypically identical alleles were identified and used interchangeably. To determine when a defect in liver formation can first be detected in *prt* mutant embryos, we examined the expression of several established liver markers⁵. The earliest genes expressed in the forming liver anlage are *hhex* and *prox1* at about 22 h.p.f.. The expression of these genes was either absent or detected only at very low levels in *prt* mutant embryos at 24 and 28 h.p.f., respectively

(Fig. 1d–i; data not shown), indicating that hepatic fate specification is defective in *prt* mutant embryos. Even though *prt* mutant embryos show no hepatic tissue at early stages, some of them start forming a liver at later stages and survive to adulthood. Indeed, whereas a liver bud can clearly be detected at 28 h.p.f. in wild-type siblings (Fig. 1g), in *prt* mutant embryos liver tissue is morphologically not distinguishable until about 50 h.p.f., when a small protrusion into the liver-forming territory becomes noticeable in a majority of them (Fig. 1c). Furthermore, differentiation markers, such as *ceruloplasmin* (*cp*), which is normally detected in the developing liver and hepatic duct, and *selenoproteinPb* (*sePb*), which is expressed only in the differentiating hepatocytes, were absent from *prt* mutant embryos at 52 h.p.f. (Fig. 1j–o). Intriguingly, this mutant displays no other morphological phenotypes. Taken together, our observations, although they cannot formally exclude the possibility that *hhex*-negative and *Prox1*-negative hepatoblasts form in the absence of *Prt* function, indicate that *prt* might be specifically involved in early liver development, namely the formation of hepatoblasts—the common progenitor of hepatocytes and bile duct cells.

To determine where the *prt* gene product is required for liver formation, we performed mosaic analysis experiments. After transplantation of rhodamine-labelled wild-type cells into homozygous mutant embryos, we found that wild-type cells contributing exclusively to the endoderm in the liver-forming region were unable to rescue the *prt* liver phenotype ($n = 4$; Fig. 2a–c). However, the presence of wild-type cells only in the lateral plate mesoderm (LPM), or in both the LPM and endoderm, led to the formation of wild-type-sized livers ($n = 1$ and $n = 3$, respectively; Fig. 2d–f). These data indicate that the *prt* gene product acts cell non-autonomously in liver formation and is required in the LPM adjacent to the forming hepatic primordium. Our findings constitute the first evidence *in vivo* that signals from the mesoderm specify the liver, in agreement with classic embryological experiments performed in chick and mouse, in which sophisticated tissue explant and recombination experiments have indicated a possible role for the cardiac and septum transversum mesoderm in liver formation^{6–9}. Thus, our data indicate that *prt* is required in the LPM to mediate the crosstalk between mesoderm and endoderm in liver formation.

To improve our understanding of the molecular nature of the *prt* phenotype, we isolated the gene disrupted by the *prt* mutation. Using a standard set of simple sequence length polymorphism (SSLP) markers, we placed the *prt* locus to linkage group 8 and through fine mapping positioned *prt* within a genomic interval of about 110 kilobases (Fig. 3a), 54.1 centimorgans from the top of the linkage group. Genome sequence analysis revealed two open reading frames within this region, that encoding F-actin-capping protein, α subunit and a *wnt2b* gene that we named *wnt2bb*, because a second *wnt2b*

¹Department of Biochemistry and Biophysics, Programs in Developmental Biology, Genetics and Human Genetics, and the Liver Center, University of California, San Francisco, 1550 Fourth Street, San Francisco, California 94143, USA. †Present addresses: National Institute for Medical Research, Division of Developmental Biology, The Ridgeway, Mill Hill, London NW7 1AA, UK (E.A.O.); Ludwig Institute for Cancer Research, Post Office Box 2008, Royal Melbourne Hospital, Victoria 3050, Australia (H.V.); Novartis Institutes for BioMedical Research, Inc., 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA (H.A.F.).

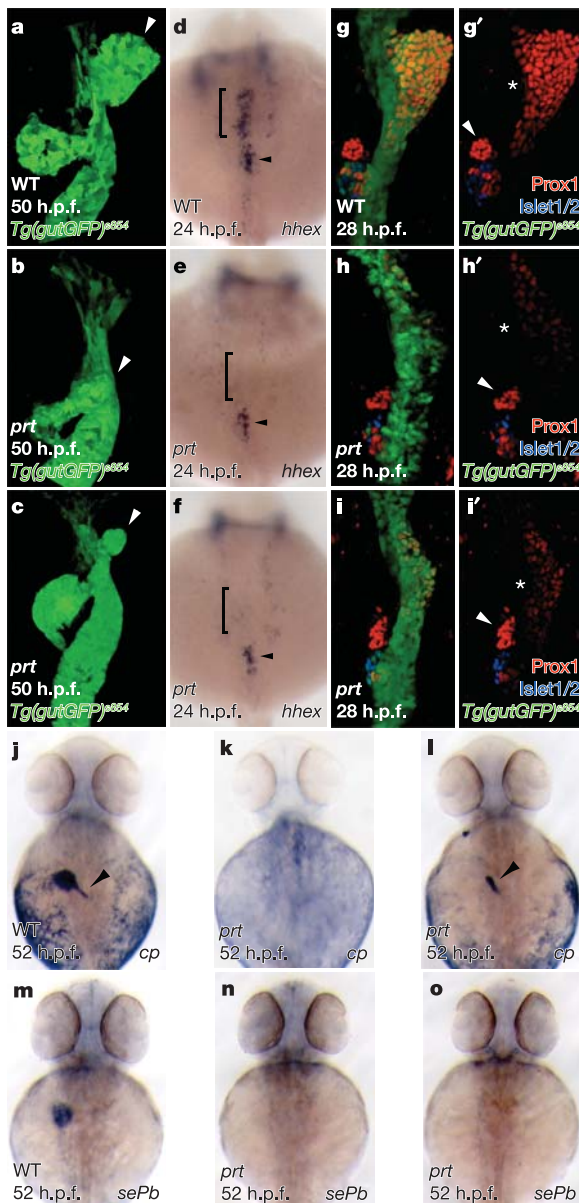


Figure 1 | Formation of hepatoblasts and differentiating hepatocytes is specifically affected in *prt* mutant embryos. **a–c**, Confocal projections of endodermal organs at 50 h.p.f. in wild-type (WT; **a**) and *prt* mutant embryos (**b**, **c**) as observed in the *Tg(gutGFP)*^{s854} line. At this stage, the *prt* mutant phenotype varies between the complete absence of a liver bud (**b**, arrowhead) and small protrusions into the liver forming area (**c**, arrowhead; 93%, *n* = 98). **d–f**, At 24 h.p.f., in contrast to the wild type (**d**), no *hhhex* expression (**e**) or expression in only a few cells (**f**) can be detected in the prospective liver anlage (brackets) in *prt* mutant embryos (**e**, 44%, *n* = 17; **f**, 56%, *n* = 22, respectively), whereas other expression domains seem unaffected (**d–f**, arrowhead). **g–i**, **g'–i'**, At 28 h.p.f., confocal projections of the *Tg(gutGFP)*^{s854} line (green) in conjunction with staining with anti-Prox1 (red) and anti-Islet1/2 (blue) antibodies reveal Prox1 expression marking the liver bud in wild-type embryos (**g'**, asterisk), whereas in *prt* mutant embryos Prox1 expression in the hepatic area is strongly decreased (**h'**, **i'**, asterisk, *n* = 10); the development of the adjacent Prox1 expressing interrenal primordium (**g'–i'**, arrowhead) and pancreatic islet (blue) appear unperturbed. **g'–i'** correspond to **g–i** with *Tg(gutGFP)*^{s854} removed for better visualization of the hepatic domain. **j–o**, In 52 h.p.f. wild-type embryos, *ceruloplasmin* (*cp*) expression is found in the forming liver and hepatic duct (**j**, arrowhead), whereas *selenoprotein P* (*sePb*) is expressed in the liver (**m**). Liver-specific expression of *cp* and *sePb* is absent from *prt* mutant embryos (**k**, **n**, **o**), although occasional expression of *cp* can be detected in what seems to be ductal tissue (**l**, arrowhead). Wild-type siblings are presented alongside two *prt* mutant embryos representing mild variability of the mutant phenotype as observed within a clutch. **a–c**, **g–i**, ventral views, anterior up; **d–f**, **j–o**, dorsal views, anterior up.

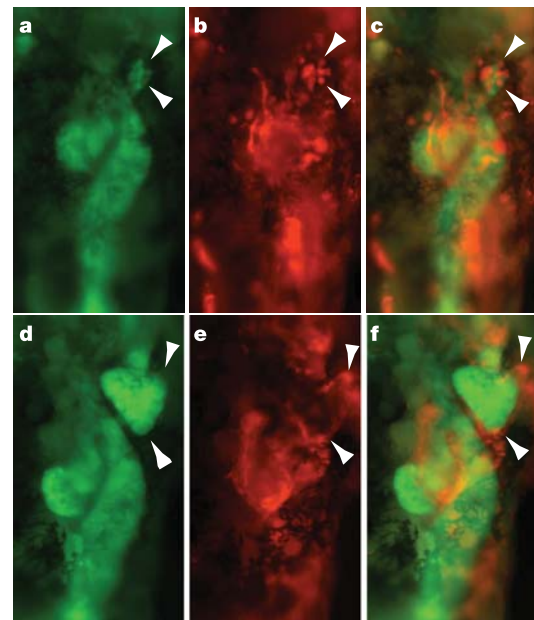


Figure 2 | Transplantation of rhodamine-labelled wild-type cells into homozygous *prt* mutant embryos reveals that *prt* functions cell-non-autonomously in promoting liver formation. Ventral views of the organ-forming region of 48 h.p.f. embryos. Green, *Tg(gutGFP)*^{s854} endodermal mutant cells; red, rhodamine-labelled wild-type cells. **a–c**, No rescue of the *prt* mutant phenotype was observed when wild-type cells contributed exclusively to the endoderm in the liver-forming area (arrowheads). **d–f**, In contrast, wild-type cells (arrowheads) in the mesenchyme surrounding the liver anlage rescued liver formation, indicating a cell-non-autonomous requirement for *prt* in liver development. **c** and **f** are merged views of **a** with **b** and **d** with **e**, respectively.

gene in zebrafish had been published previously¹⁰ and will be referred to as *wnt2ba*. This nomenclature is based on sequence and synteny analyses (Supplementary Fig. 1). Sequencing showed that all three *prt* mutant alleles contain point mutations in *wnt2bb* (Fig. 3b). The *prt*^{s403} allele bears a T→C base change at position 454, leading to the replacement of a conserved serine residue by proline. A C→T change at position 538 in the *prt*^{s404} allele creates a premature stop codon, deleting 245 amino acids from the carboxy terminus of the protein. Similarly, a T→A change at position 282 in the *prt*^{s416} allele creates a premature stop codon, truncating the protein by 303 amino acids.

To investigate further whether a decrease in Wnt2bb function causes the liver phenotype observed in *prt* mutants, we knocked down its function by using morpholino antisense oligonucleotides (MO). Injection of 2 ng of *wnt2bb* MO into *Tg(gutGFP)*^{s854} embryos led to a complete absence or strong decrease in hepatic tissue in otherwise wild-type-looking embryos at 52 h.p.f. (*n* = 107; Fig. 3c). Overall, the tight genetic linkage, the presence of severe molecular lesions and MO phenocopy indicate that *wnt2bb* is the gene affected by the *prt* mutations.

Because Wnt2b has been shown to be able to signal through β-catenin¹¹, we tested whether interfering with β-catenin signalling would affect hepatic specification, using a transgenic zebrafish line expressing an inhibitor of Wnt/β-catenin signalling, dnTcf3-GFP, under the heat-shock promoter¹². These embryos received a single heat shock at one of several different developmental stages, from 16 h.p.f. (14-somite stage) to 25 h.p.f. Expression of *hhhex* was assessed at 25 h.p.f. and of *foxA1* at 48 h.p.f. Using these markers we found that inhibiting β-catenin signalling between 16 and 21 h.p.f. resulted in the absence or strong decrease in hepatic tissue, whereas transgenic embryos heat-shocked at 25 h.p.f. did form a liver, although of varying size (Supplementary Fig. 2). These data

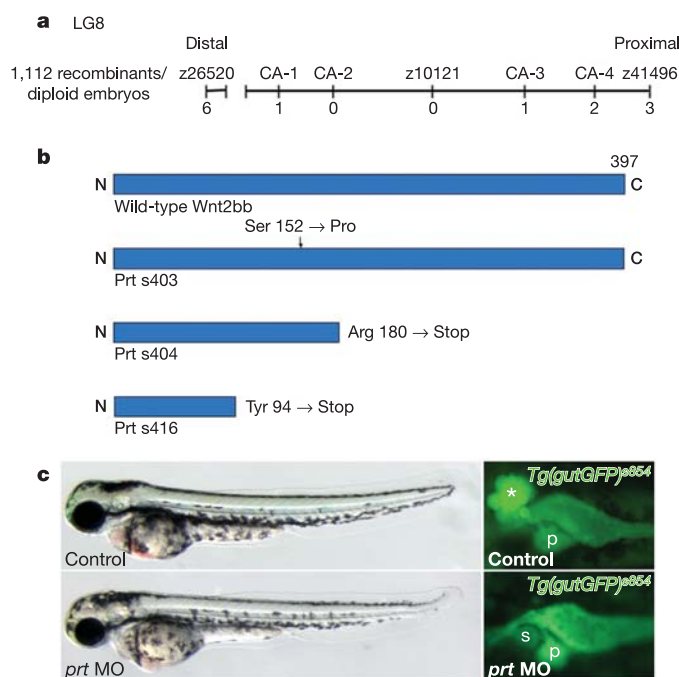


Figure 3 | *prt* encodes Wnt2bb. **a**, Genetic map of the *prt* region. Numbers below SLP markers indicate recombination events. **b**, Alterations to Wnt2bb in the three *prt* mutant alleles. **c**, Knock-down of *wnt2bb* by the injection of 2 ng of MO into the *Tg(gutGFP)^{s854}* line led to a complete absence of liver (asterisk in control), whereas pancreas (p), swim bladder (s) and the rest of the embryo seemed unaffected at 54 h.p.f. Brightfield pictures show lateral views, pictures of *Tg(gutGFP)^{s854}* ventral views; anterior left.

implicate β -catenin in early liver development; together with the data shown above they indicate that Wnt2bb might regulate hepatic specification through the canonical Wnt signalling pathway. In addition, we found that injection of *prt* messenger RNA into the early embryo led to an increase in dorsal fates (Supplementary Fig. 2), further indicating that Wnt2bb can function as a canonical Wnt.

In the developing zebrafish embryo, *wnt2bb* is expressed in a dynamic pattern (Fig. 4a–h) that is similar to that reported for the mouse and chick orthologues^{13,14}. Transcripts can be first detected at the onset of somitogenesis in the posterior LPM, predominantly in cells fated to give rise to the endothelial and blood lineages (data not shown). At 18 and 21 h.p.f. (18-somite and 24-somite stages, respectively), *wnt2bb* is expressed in discrete bilateral domains close to the midline in the liver-forming region, as well as in a subset of branchial arch tissue and the developing vasculature (Fig. 4a–d). Because the vascular expression obscures the other expression domains, we investigated *wnt2bb* expression in *cloche* (*clo*) mutant embryos, which are devoid of endothelial cells. As expected, no vascular expression can be found in *clo* mutants, but the other domains of expression become clearly visible (Fig. 4e), revealing the bilateral expression in the liver-forming region. (Nascent endothelial cells have been implicated in early mouse liver formation¹⁵, promoting outgrowth after hepatic specification. However, in *clo* mutant embryos early liver development occurs unperturbed³, probably excluding a role for vascular *wnt2bb* expression in liver specification.) By 24 h.p.f., *wnt2bb* expression has disappeared from endothelial cells, whereas expression in the organ-forming region is still present (Fig. 4f). To determine whether *wnt2bb* is expressed in the LPM adjacent to the liver or in the organ-forming endoderm, we examined *wnt2bb* expression in *casanova* (*cas*) mutant embryos, which lack all endodermal cells. At 24 h.p.f., bilateral *wnt2bb* expression in the liver-forming region is present in *cas* mutants, although at a variable distance from the midline, which is probably

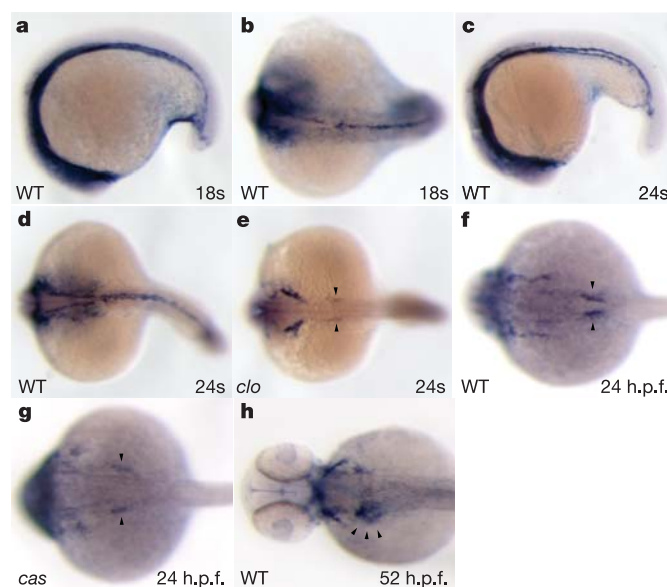


Figure 4 | *wnt2bb* expression is dynamic during organogenesis stages.

a, b, At the 18-somite (18 s) stage (18 h.p.f.), *wnt2bb* is expressed in the developing vasculature, shown in lateral (**a**) and dorsal views (**b**). Similarly, at the 24-somite stage the vascular expression (**c, d**) is obscuring overlapping expression domains (**e**). **e**, Starting at the 18-somite stage (data not shown), but more distinctly at the 24-somite stage (21 h.p.f.), *wnt2bb* is expressed in bilateral domains of the LPM at the level of the prospective liver-forming area; 18–24-somite stages in zebrafish correspond to 7.5–8.5 days after conception in mouse. **e–h**, Analysis of *cloche* (**e**) and *casanova* (**g**) mutant embryos confirmed the mesodermal nature of the bilateral expression domains of *wnt2bb* (arrowheads) neighbouring the future liver anlage. **f**, At 24 h.p.f., *wnt2bb* expression persists in the bilateral LPM domains (arrowheads) and is also present in the endocardium, branchial arch anlagen, blood island and parts of the eyes and otic vesicles. **h**, At 52 h.p.f. *wnt2bb* expression is present in the atrioventricular valve, branchial arches, eye, ear and the mesenchyme surrounding the liver (arrowheads). **a, c**, Lateral views, anterior left; **b, d–i**, dorsal views, anterior left.

due to defects in LPM migration to the midline in the absence of endoderm (Fig. 4g). At 52 h.p.f., *wnt2bb* expression is still present in the mesenchyme adjacent to the liver (Fig. 4h). These findings show that the bilateral expression of *wnt2bb* in the pre-hepatic region is in the LPM and that the highly restricted pattern of *wnt2bb* expression in the LPM does not depend on signals from the endoderm. Thus, the spatiotemporal expression of *wnt2bb* concurs with its requirement in the LPM of the liver-forming region, and also with the time window of liver specification determined by inhibiting β -catenin signalling.

Thus, we have presented genetic evidence for a positive role for Wnt signalling in liver formation, distinct from its well-described roles in anterior endoderm specification¹⁶ and in liver growth and regeneration^{17–19}. Our data indicate a possible model in which *wnt2bb* expressed bilaterally in the LPM specifies a single hepatic anlage in the adjacent competent endoderm³. Moreover, Prt/Wnt2bb function might also regulate hepatoblast proliferation directly or indirectly through factors such as Hhex²⁰ and Prox1 (ref. 21). Although the expression of the gene encoding the Wnt antagonist Secreted frizzled-related protein 5 (sFRP5) in the foregut endoderm giving rise to the liver in mouse and *Xenopus*^{22,23} has so far been interpreted as negative regulation of hepatic development by Wnt signalling¹, we propose that its presence is required to modulate Wnt signalling, thereby allowing further development of the liver. Furthermore, our findings indicate that the asymmetric outgrowth of the liver occurs independently of its specification, possibly in conjunction with the looping of the gut tube and asymmetric placement of the LPM²⁴.

This work, in concert with that by others^{6–9,25–27}, indicates that the crosstalk between mesoderm and endoderm might represent a common theme in establishing endodermal organ fates along the gut tube. The highly restricted expression domain of *prr/wnt2bb* along the anteroposterior axis of the LPM is reminiscent of the defined expression patterns of *Wnt* genes along the oral–aboral axis of the sea anemone *N. vectensis*²⁸ and the chicken intestinal tract^{29,30}. On the basis of those studies and the function and expression pattern of *prr/wnt2bb* described here, we propose that *Wnt* genes are key spatial and temporal determinants for the induction of accessory organs along the gut tube.

METHODS

Zebrafish strains. Embryos and adult fish were raised and maintained under standard laboratory conditions. We used the following lines: *prr*^{s403}, *prr*^{s404}, *prr*^{s416}, *clo*^{s5}, *cas*^{tu56}, *Tg(gutGFP)*^{s854} and *Tg(hsΔTcf3-GFP)*.

Mapping. We mapped the mutation to linkage group 8 using a set of SSLP markers. For fine mapping, 1,112 mutant embryos were tested with SSLP markers in the critical interval (Fig. 2a). *prr/wnt2bb* complementary DNA was isolated, sequenced and analysed from *Tg(gutGFP)*^{s854} and the three mutant alleles.

Mosaic experiments. Donor embryos were injected with 2.5% rhodamine-dextran, and cells transplanted into hosts at the 1,000-cell stage. To target wild-type cells to the mesoderm, cells were transplanted to the margin. Two independent groups of cells were transplanted into each host embryo, 90° apart, to increase the likelihood of contributing to the LPM.

Histochemical methods. Whole-mount *in situ* hybridization was performed as previously described^{5,24}, using the following probes: *hhex*, *cp*, *sePb*, *gsc* and *foxA1*. For generating *prr* *in situ* probe, 850 base pairs at the 3' end of the gene were cloned into the pCS2+ vector.

We used the following antibodies: polyclonal antibody against Prox1 (rabbit, 1:1000; Chemicon), monoclonal antibody against Isl1/2 (1:15; Developmental Studies Hybridoma Bank, clone 39.4D5) and fluorescently conjugated antibodies from Molecular Probes and Jackson Laboratories. Embryos were fixed with 2% PFA overnight at 4°C. Manual removal of the yolk was followed by 60 min incubation with DNaseI at 37°C, followed by whole-mount antibody staining in PBST (1% Triton in PBS pH 7.3) with 5% sheep serum and 1% DMSO added. Stained embryos were embedded in low-melting-point agarose and cut into 200-μm sections with a vibratome. Images were acquired using confocal microscopy.

Morpholino antisense oligonucleotide injection and mRNA overexpression. We used a morpholino oligonucleotide targeted against an ATG upstream of the translational start site of *prr* with the following sequence: 5'-GTGTGCCA TATAAAGTATTCCCG-3'. *Tg(gutGFP)*^{s854} embryos were injected at the one-cell stage with ~2 ng of morpholino and assayed between 48–54 h.p.f. For mRNA overexpression, one-cell-stage embryos were injected with 50–150 pg of *prr* mRNA.

Additional details on materials and methods as well as information regarding the *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis and screen can be found in the Supplementary Information.

Received 31 March; accepted 15 May 2006.

Published online 21 June 2006.

- Lemaigre, F. & Zaret, K. S. Liver development update: new embryo models, cell lineage control, and morphogenesis. *Curr. Opin. Genet. Dev.* **14**, 582–590 (2004).
- Zhao, R. & Duncan, S. A. Embryonic development of the liver. *Hepatology* **41**, 956–967 (2005).
- Lee, C. S., Friedman, J. R., Fulmer, J. T. & Kaestner, K. H. The initiation of liver development is dependent on Foxa transcription factors. *Nature* **435**, 944–947 (2005).
- Holtzinger, A. & Evans, T. Gata4 regulates the formation of multiple organs. *Development* **132**, 4005–4014 (2005).
- Field, H. A., Ober, E. A., Roeser, T. & Stainier, D. Y. Formation of the digestive system in zebrafish. I. Liver morphogenesis. *Dev. Biol.* **253**, 279–290 (2003).
- Le Douarin, N. M. An experimental analysis of liver development. *Med. Biol.* **53**, 427–455 (1975).
- Fukuda-Taira, S. Hepatic induction in the avian embryo: specificity of reactive endoderm and inductive mesoderm. *J. Embryol. Exp. Morphol.* **63**, 111–125 (1981).
- Jung, J., Zheng, M., Goldfarb, M. & Zaret, K. S. Initiation of mammalian liver

development from endoderm by fibroblast growth factors. *Science* **284**, 1998–2003 (1999).

- Rossi, J. M., Dunn, N. R., Hogan, B. L. & Zaret, K. S. Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. *Genes Dev.* **15**, 1998–2009 (2001).
- Ng, J. K. *et al.* The limb identity gene Tbx5 promotes limb initiation by interacting with Wnt2b and Fgf10. *Development* **129**, 5161–5170 (2002).
- Landesman, Y. & Sokol, S. Y. Xwnt-2b is a novel axis-inducing *Xenopus* Wnt, which is expressed in embryonic brain. *Mech. Dev.* **63**, 199–209 (1997).
- Lewis, J. L. *et al.* Reiterated Wnt signaling during zebrafish neural crest development. *Development* **131**, 1299–1308 (2004).
- Jasoni, C., Hendrickson, A. & Roelink, H. Analysis of chicken Wnt-13 expression demonstrates coincidence with cell division in the developing eye and is consistent with a role in induction. *Dev. Dyn.* **215**, 215–224 (1999).
- Zakin, L. D. *et al.* Structure and expression of Wnt13, a novel mouse Wnt2 related gene. *Mech. Dev.* **73**, 107–116 (1998).
- Matsumoto, K., Yoshitomi, H., Rossant, J. & Zaret, K. S. Liver organogenesis promoted by endothelial cells prior to vascular function. *Science* **294**, 559–563 (2001).
- Zorn, A. M., Butler, K. & Gurdon, J. B. Anterior endomesoderm specification in *Xenopus* by Wnt/β-catenin and TGF-β signalling pathways. *Dev. Biol.* **209**, 282–297 (1999).
- Suksaweang, S. *et al.* Morphogenesis of chicken liver: identification of localized growth zones and the role of β-catenin/Wnt in size regulation. *Dev. Biol.* **266**, 109–122 (2004).
- Hussain, S. Z. *et al.* Wnt impacts growth and differentiation in ex vivo liver development. *Exp. Cell Res.* **292**, 157–169 (2004).
- Monga, S. P., Padiaditakis, P., Mule, K., Stolz, D. B. & Michalopoulos, G. K. Changes in WNT/β-catenin pathway during regulated growth in rat liver regeneration. *Hepatology* **33**, 1098–1109 (2001).
- Bort, R., Martinez-Barbera, J. P., Beddington, R. S. & Zaret, K. S. Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. *Development* **131**, 797–806 (2004).
- Sosa-Pineda, B., Wigle, J. T. & Oliver, G. Hepatocyte migration during liver development requires Prox1. *Nature Genet.* **25**, 254–255 (2000).
- Finley, K. R., Tennessen, J. & Shawlot, W. The mouse secreted frizzled-related protein 5 gene is expressed in the anterior visceral endoderm and foregut endoderm during early post-implantation development. *Gene Expr. Patterns* **3**, 681–684 (2003).
- Pilcher, K. E. & Krieg, P. A. Expression of the Wnt inhibitor, sFRP5, in the gut endoderm of *Xenopus*. *Gene Expr. Patterns* **2**, 369–372 (2002).
- Horne-Badovinac, S., Rebagliati, M. & Stainier, D. Y. A cellular framework for gut-looping morphogenesis in zebrafish. *Science* **302**, 662–665 (2003).
- Roberts, D. J., Smith, D. M., Goff, D. J. & Tabin, C. J. Epithelial-mesenchymal signaling during the regionalization of the chick gut. *Development* **125**, 2791–2801 (1998).
- Wells, J. M. & Melton, D. A. Early mouse endoderm is patterned by soluble factors from adjacent germ layers. *Development* **127**, 1563–1572 (2000).
- Kumar, M., Jordan, N., Melton, D. & Grapin-Botton, A. Signals from lateral plate mesoderm instruct endoderm toward a pancreatic fate. *Dev. Biol.* **259**, 109–122 (2003).
- Kusserow, A. *et al.* Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* **433**, 156–160 (2005).
- McBride, H. J., Fatke, B. & Fraser, S. E. Wnt signaling components in the chicken intestinal tract. *Dev. Biol.* **256**, 18–33 (2003).
- Theodosiou, N. A. & Tabin, C. J. Wnt signaling during development of the gastrointestinal tract. *Dev. Biol.* **259**, 258–271 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all the members of the Stainier laboratory for their support during the screen; A. Ayala, S. Waldron and N. Zvenigorodsky for expert help with the fish; L. Trinh, S.-W. Jin, H. Knaut and J.-P. Vincent for discussions and critical readings of the manuscript; S. Pleasure for providing a Prox1 antibody before commercialization; and R. Dorsky for providing *Tg(hsΔTcf3-GFP)*. E.A.O. was supported by the UCSF Liver Center through an NIH pilot feasibility grant, and H.V. by a long-term fellowship of the Human Frontier Science Program. This work was supported in part by grants from the NIH (NIDDK) and the Packard Foundation to D.Y.R.S.

Author Information The *prr/wnt2bb* sequence is deposited in GenBank under accession number DQ231559. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.A.O. (eober@nimr.mrc.ac.uk) or D.Y.R.S. (didier_stainier@biochem.ucsf.edu).

LETTERS

Microstimulation of inferotemporal cortex influences face categorization

Seyed-Reza Afraz^{1,2}, Roozbeh Kiani^{1,2} & Hossein Esteky¹⁻³

The inferior temporal cortex (IT) of primates is thought to be the final visual area in the ventral stream of cortical areas responsible for object recognition^{1,2}. Consistent with this hypothesis, single IT neurons respond selectively to highly complex visual stimuli such as faces³⁻⁶. However, a direct causal link between the activity of face-selective neurons and face perception has not been demonstrated. In the present study of macaque monkeys, we artificially activated small clusters of IT neurons by means of electrical microstimulation while the monkeys performed a categorization task, judging whether noisy visual images belonged to 'face' or 'non-face' categories. Here we show that microstimulation of face-selective sites, but not other sites, strongly biased the monkeys' decisions towards the face category. The magnitude of the effect depended upon the degree of face selectivity of the stimulation site, the size of the stimulated cluster of face-selective neurons, and the exact timing of microstimulation. Our results establish a causal relationship between the activity of face-selective neurons and face perception.

We trained two adult macaque monkeys to perform a face/non-face categorization task upon viewing single images from one or the other category that were systematically degraded by varying amounts of visual signal. We chose the various signal levels to create a range of difficulties spanning a psychophysical threshold: categorization was easy in some trials and difficult in others (Fig. 1a). In each trial, the monkey was presented briefly (54 ms) with a face or a non-face image degraded by noise. Subsequently, the monkey was required to make a saccadic eye movement to one of two targets to indicate whether the image was a face or a non-face. Each correct response was rewarded by a drop of juice. For pure noise stimuli (Fig. 1a, '0% visual signal'), the monkey was rewarded randomly with a probability of 0.5.

Our central experimental question was whether electrical microstimulation of clusters of face-selective IT neurons would bias the monkeys' choices towards the face category. Because of its relatively precise temporal and spatial characteristics, microstimulation is a particularly powerful tool for establishing causal relationships between physiologically characterized neurons and behavioural performance⁷⁻¹⁰. Even weak microstimulation pulses excite many neurons simultaneously¹¹⁻¹³; successful use of extracellular microstimulation therefore relies on structural regularities within the cortex, such as the presence of cortical columns^{14,15}. Face-selective neurons are found in relatively large clusters in IT¹⁶⁻¹⁸, making them an optimal target for microstimulation.

In each experimental session, we assessed the face selectivity of multiunit clusters of neurons at regular intervals (minimum steps of 150 μm) through a single electrode penetration in IT cortex. At each recording site, selectivity was determined by presenting a large number of face and non-face images while the monkey passively fixated a small fixation point on the monitor screen. Face/non-face stimulus selectivity of multiunit responses was quantified with the d'

index (see Methods). A d' value of zero indicates indistinguishable responses to faces and non-faces. Increasingly positive d' values indicate progressively better selectivity for faces.

After recording from several sites within a track (mean number of recorded sites in each track was 4), the electrode was positioned in between the recorded sites and neural response selectivity was determined again. Altogether, we assessed stimulus selectivity at 348 recording sites in 86 electrode penetrations in two monkeys (46 and 40 in monkeys FR and KH, respectively). We conducted microstimulation experiments at 31 face-selective sites and 55 non-selective sites, while the monkey performed the object categorization task. Selectivity for faces was defined as having a d' value >1 .

Microstimulation consisted of bipolar current pulses of 50 μA delivered at 200 Hz (refs 19, 20). The stimulation pulses were biphasic, with the cathodal pulse leading. Each pulse was 0.2 ms in duration with 0.1 ms between the cathodal and anodal phase. Each experiment contained three microstimulation conditions differing in the exact time of stimulation delivery as well as an un-stimulated control condition (Fig. 1b). Stimulating pulses were delivered for 50 ms in one of three time periods following onset of the visual stimulus: 0–50 ms, 50–100 ms or 100–150 ms. The first period was

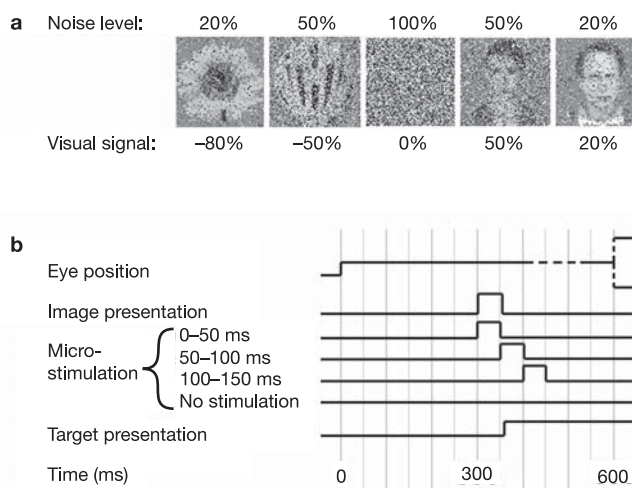


Figure 1 | Visual stimuli and event timing. In each experimental session, the neural stimulus selectivity of several neighbouring cortical sites was first determined in a fixation task using luminance-matched face and non-face greyscale images. Then, in the second part of the experiment (**a**), face and non-face images with varying amounts of noise were used in a face categorization task. **b**, Timing of events in each categorization trial. One of the four possible microstimulation conditions shown was applied randomly in each trial.

¹School of Cognitive Sciences, Institute for Studies in Theoretical Physics and Mathematics, Tehran, 19395, Iran. ²Research Center for Brain and Cognitive Sciences,

³Neuroscience Research Center, Shaheed Beheshti University of Medical Sciences, Tehran, 19385, Iran.

before the earliest visual responses normally observed in IT, and the last two periods correspond to the earliest and later IT responses, respectively^{21–23}. The three stimulation conditions and the control trials were randomly interleaved in each experiment.

To reveal the impact of microstimulation on behaviour, the monkey's performance in the categorization task was plotted as the proportion of 'face' choices as a function of the visual stimulus signal for face and non-face images (Fig. 2). We used positive visual signal values for faces and negative values for non-faces to create a continuum. Logistic regression analysis was used to determine whether the three microstimulation time conditions caused a significant shift in the psychometric functions compared to the non-stimulated condition.

Figure 2 illustrates results obtained in two typical microstimulation experiments. In both experiments, microstimulation during the 50–100 ms interval shifted the monkeys' choices significantly towards the face category (Fig. 2a, logistic regression, $P < 0.001$; Fig. 2b, $P < 0.01$). Microstimulation during the 100–150 ms interval

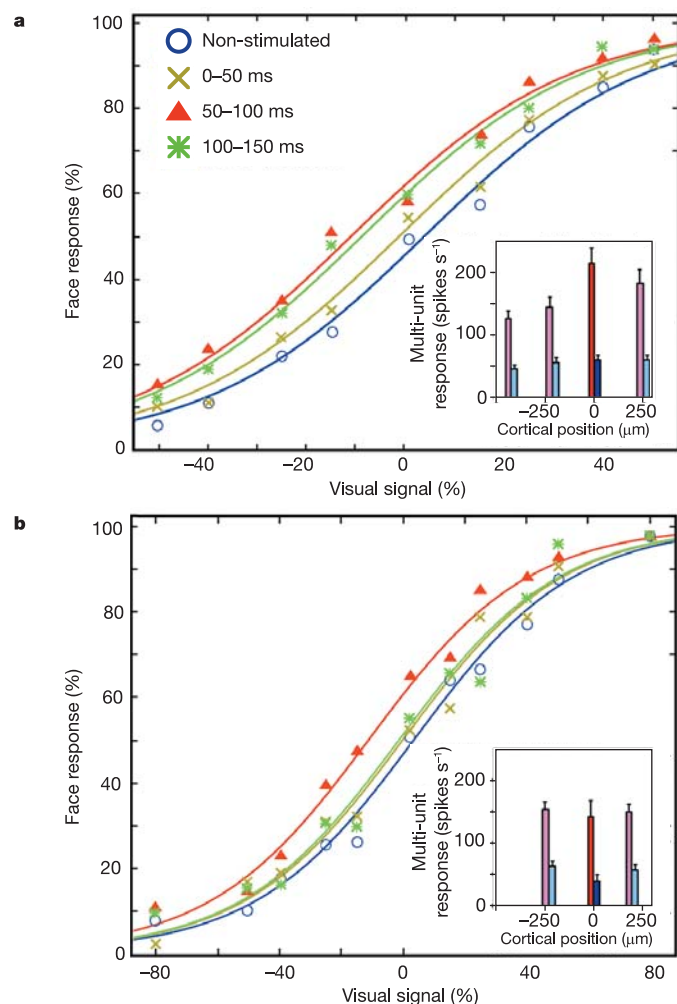


Figure 2 | Effect of microstimulation of two representative face-selective neural clusters in IT cortex. **a**, Monkey KH; **b**, monkey FR. Data points show the proportion of face choices for different levels of visual signal in the images for different microstimulation conditions. The curves are logistic regression fits to the data points. The insets show averaged multiunit responses of the corresponding stimulated sites and their neighbouring sites. The inset abscissa shows the cortical position of the electrode tip along the recording track relative to the stimulated site (zero on the abscissa). The inset ordinate shows the averaged multi-unit neural responses. Responses to face and non-face stimuli are represented by red and blue bars, respectively. Colours are highlighted for the stimulated site. Error bars, s.e.m.

biased choices significantly in the experiment of Fig. 2a ($P < 0.001$), but not in the other experimental session depicted in Fig. 2b ($P = 0.283$). (See Supplementary Fig. 1 for more examples of psychometric shifts.) Microstimulation resulted in significant shift of the psychometric function in favour of face choices in at least one of the stimulation conditions for 19 of 31 face selective sites (61%; 9 in right hemisphere and 10 in left hemisphere) and one non-face site with $d' = 0.94$ (Fig. 3). No significant shift in favour of non-face choices was ever observed.

The impact of microstimulation on perceptual decisions increased as a function of the neural face selectivity of the stimulated site. The scatter plots of Fig. 3a show the correlation between the degree of face selectivity of the stimulated sites and the shift of the psychometric function in different microstimulation conditions. The strongest correlation was found for microstimulation at 50–100 ms after image onset ($r = 0.643$, $P < 0.0001$). Microstimulation at 100–150 ms also showed a significant correlation ($r = 0.539$, $P < 0.0001$). No significant correlation was observed for 0–50 ms ($r = 0.147$, $P = 0.18$).

It has previously been shown that following psychometric function shift the amount of total reward gained by the animal drops in proportion to the amount of shift⁷. Consistently, we observed significant correlations between reward loss and shift value for 50–100 ms and 100–150 ms conditions ($r = 0.42$, $P < 0.001$ and $r = 0.47$, $P < 0.001$, respectively) but not for the 0–50 ms condition ($r = -0.02$, $P = 0.83$).

The impact of microstimulation on perceptual decisions was much larger when current was injected into larger clusters of face-selective neurons. Recall that we measured stimulus selectivity at recording sites adjacent to the stimulation site as well as at the stimulation site itself. When adjacent sites exhibit selectivity similar to that of the recorded site, we may infer that the cluster of physiologically homogeneous neurons is larger, at least along the dimension of our electrode track.

Figure 4a summarizes the effect of stimulating clusters of different sizes. Stimulation effects were substantially more pronounced for larger clusters of face-selective neurons (Fig. 4a; black bars) as compared to smaller clusters (Fig. 4a; grey bars). On average, there was no significant shift in the psychometric function following microstimulation of cortical clusters lacking face selectivity (Fig. 4b). A two-way analysis of variance showed a significant effect of both neighbourhood ($F(1,87) = 11.248$, $P = 0.001$) and stimulation timing ($F(2,87) = 6.092$, $P = 0.003$) on the averaged values of the psychometric function shifts across all face-selective sites (sites with $d' > 1$). No such significant effect was observed in non-selective sites. The averaged d' of neighbouring sites was correlated with the effect of microstimulation: the correlating coefficients for 0–50 ms, 50–100 ms and 100–150 ms stimulation conditions are $r = 0.12$, $P = 0.31$; $r = 0.49$, $P < 0.001$; and $r = 0.44$, $P < 0.001$, respectively.

To prevent the monkeys memorizing specific examples, we used a large image bank, thereby making it unlikely that the monkeys could memorize the specific images. To further examine the unlikely event of whether the monkeys simply memorized all the images in the image set, a behavioural experiment was conducted after the completion of the training in each monkey. In these experiments, 40 novel images (20 faces and 20 non-face objects) were intermixed with 40 familiar face and non-face images (randomly chosen from the learned image bank). The stimuli were presented to the monkey without any visual noise. The face/non-face discrimination performance of the monkeys was measured in several behavioural sessions. In each session, we used a new set of novel images. The monkeys' performance for novel stimuli was as good as it was for familiar stimuli (both above 95%) from the very beginning of the behavioural sessions. Furthermore, to prevent the monkeys using a general rule other than face/non-face categorization, such as detection of living versus non-living objects, we included non-face animate object

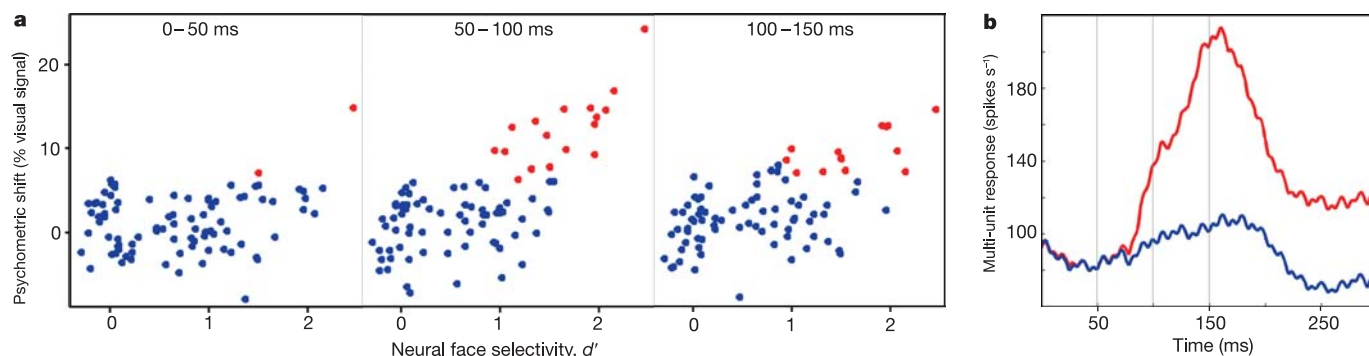


Figure 3 | Correlation between face selectivity of stimulated sites and the behavioural impact of microstimulation. **a**, Positive values on the y-axis represent psychometric function shifts in favour of face choices. Large d' values indicate higher selectivity for faces. Red data points indicate statistically significant shifts of the psychometric function. The correlation is significant for microstimulation at 50–100 ms and 100–150 ms conditions,

but not for 0–50 ms (see text for details). **b**, Histogram of multiunit neural responses (smoothed with a 20-ms sliding window) to face (red line) and non-face (blue line) stimuli averaged from all stimulated face-selective ($d' > 1$) sites. Different microstimulation time windows are depicted by vertical lines.

images (for example, human and animal bodies from a back view or with the face cut out) in our non-face set and included artefact faces (for example, masks and sculpture faces) in the face set. The possibility of familiarity being a factor was also controlled by using images of human and monkey bodies (which are presumably as familiar objects as faces) in the non-face set.

Our findings demonstrate a causal relationship between IT neural activity and visual object perception and categorization. Although a general role for IT cortex in object perception has been demonstrated previously in cortical ablation experiments^{24,25}, our data extend causality to a much finer spatial scale. In addition, our data demonstrate that single neuron response properties provide important clues to the functional role of neurons in perception, even for highly complex stimuli such as faces. The functional role of face-selective neurons in behaviour has been debated, but our data clearly show that this role includes, at the very least, categorization of objects into faces and non-faces.

The 50–100 ms stimulation period overlaps with the earliest time window shown to convey information about stimulus category^{22,23,26}. The significant effect of microstimulation in this period suggests that downstream cortical areas can use such early information, in addition to the later 100–150 ms neural activity, for decision making.

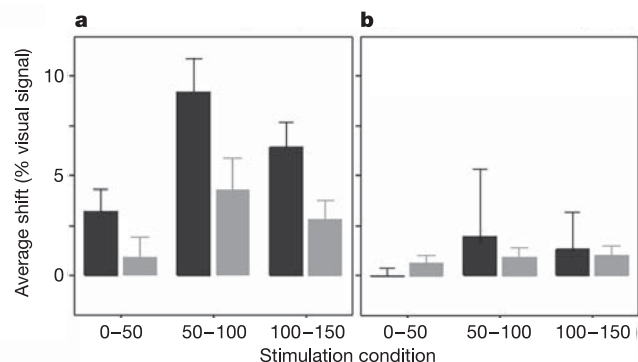


Figure 4 | Effect of stimulus selectivity of neighbouring cortical sites on microstimulation results. Averaged shift in the psychometric function for the three stimulation conditions is shown for all stimulated face-selective (**a**) and non-selective (**b**) sites. Face-selective and non-selective sites were defined by $d' > 1$ and $d' \leq 1$, respectively. Black columns represent sites with face-selective neighbours in their vicinity ($\pm 500 \mu\text{m}$), and grey columns show sites with non-selective neighbour(s). Positive numbers on the y-axis show shifts in favour of face choices. Error bars, s.e.m.

METHODS

Behavioural tasks. Each session started with a passive fixation task in which macaque monkeys (*Macaca mulatta*) were required to maintain fixation in a $4^\circ \times 4^\circ$ window at the centre of the screen. Following 300 ms of fixation, a sequence of visual stimuli ($7^\circ \times 7^\circ$ in size) was presented to the monkey. Each image was presented for 200 ms without blank intervals between images^{23,27,28}, 7–10 times pseudorandomly. The images were greyscale photographs of 30 face objects and 60 non-face objects chosen randomly from a bank of 600 images.

In the second phase of the experiment, monkeys performed a face/non-face categorization task. The monkey started a trial by fixating on the fixation spot for 300 ms. Then a noisy image was presented for 54 ms, followed immediately by two small response targets presented 10° to the left and right of the screen centre. The left and right targets represented face and non-face responses, respectively. The monkey was required to make a saccade to the correct target no later than 660 ms after the onset of targets.

To minimize the monkey's behavioural choice bias, we used a correction scheme⁷. The monkey entered a set of correction trials if it made three consecutive errors within a single category (face or non-face). Upon entering a correction trial, images from the neglected category were presented until the monkey chose that category correctly. All data collected from correction trials were discarded from the analysis. The monkey entered a correction trial in 32 of the 86 sessions, resulting in exclusion of 5.7% of trials in those sessions.

In the categorization task, 11 or 9 signal levels were used in monkey FR and 9 signal levels were used in monkey KH. Each signal level was generated by assigning a uniformly distributed greyscale value to $X\%$ of image pixels, where X is the absolute signal level (positive for faces and negative for non-faces). Noisy face and non-face images create a continuum of task relevant visual signal extending from noiseless faces (100) to full noise images (0) to noiseless non-faces (-100). For each signal level, 16 face and 16 non-face images were randomly selected from the image bank.

Electrophysiology. Recordings were made on an evenly spaced grid, with 1-mm intervals between penetrations over a wide region of the lower bank of STS and TEa cortices²³ (left hemisphere 14 to 21 mm anterior to interauricular line in FR, and right hemisphere 14 to 20 mm anterior to interauricular line in KH). The recording positions were determined stereotactically by referring to magnetic resonance images acquired before the surgery. Multiunit neural responses were recorded through tungsten microelectrodes ($0.4\text{--}1.0 \text{ M}\Omega$). Neural selectivity of neighbouring sites within $\pm 500 \mu\text{m}$ from the stimulated site along each recording track was determined as the electrode was advanced. The recorded positions were separated by at least $150 \mu\text{m}$ (mean, $296 \mu\text{m}$). After determining the neighbourhood selectivity, the electrode tip was positioned in the middle of the recorded area and remained there through the rest of the experiment. The neural selectivity in this site was verified before starting the categorization task.

Data analysis. Mean multiunit discharge for each stimulus was measured in a period 70–200 ms after the image onset. The degree of selectivity of each cortical site for face versus non-face images was measured by the d' index²⁹, based on the following formula:

$$d' = \frac{[M(f) - M(nf)]}{\sqrt{\frac{\sigma^2(f) + \sigma^2(nf)}{2}}}$$

where $M(f)$ and $M(nf)$ are the mean multiunit response to face and non-face images, respectively, and $\sigma^2(f)$ and $\sigma^2(nf)$ are the variance of the distributions of neural responses to face and non-face images, respectively.

To calculate the shift in psychometric function, logistic curves were fitted to the monkey's responses in the categorization task, based on the following formula:

$$P(x) = \frac{1}{1 + e^{-(\alpha + \beta x + \lambda_1 I_1 + \lambda_2 I_2 + \lambda_3 I_3)}}$$

where x is the visual signal and $P(x)$ is the probability of face response, I_1 , I_2 and I_3 indicate the presence or absence of microstimulation in the three periods (one for stimulation present and zero for stimulation absent condition), and α , β and λ are free parameters (with λ_1 to λ_3 indicating the three stimulation conditions respectively) that were fitted using the maximum likelihood fitting procedure³⁰. The fit was performed separately for all of the behavioural data obtained in each experimental session (86 fits). The microstimulation effect in each site was considered significant if λ_i was significantly different from zero ($P < 0.05$). The shift of the psychometric function in each stimulation condition was defined as the change in the visual signal that would have induced a behavioural effect comparable to that of the microstimulation. This is equal to λ_i/β in the logistic fit. Similar methods have been used in other microstimulation studies^{7,19}. To reduce the number of the free parameters, our logistic fit assumes a similar slope for the psychometric curves in different stimulation conditions. Allowing different slopes did not improve the fit and was not critical to the results. There was no significant correlation between shift values and slope changes after free fitting of the different conditions (see Supplementary Table 1).

Received 25 January; accepted 16 June 2006.

Published online 26 July 2006.

1. Tanaka, K. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* **19**, 109–139 (1996).
2. Logothetis, N. K. & Sheinberg, D. L. Visual object recognition. *Annu. Rev. Neurosci.* **19**, 577–621 (1996).
3. Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* **4**, 2051–2062 (1984).
4. Perrett, D. I., Rolls, E. T. & Caan, W. Visual neurones responsive to faces in the monkey temporal cortex. *Exp. Brain Res.* **47**, 329–342 (1982).
5. Rolls, E. T. Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Hum. Neurobiol.* **3**, 209–222 (1984).
6. Gross, C. G. Processing the facial image: a brief history. *Am. Psychol.* **60**, 755–763 (2005).
7. Salzman, C. D., Muraug, C. M., Britten, K. H. & Newsome, W. T. Microstimulation in visual area MT: effects on direction discrimination performance. *J. Neurosci.* **12**, 2331–2355 (1992).
8. Romo, R., Hernandez, A., Zainos, A. & Salinas, E. Somatosensory discrimination based on cortical microstimulation. *Nature* **392**, 387–390 (1998).
9. Romo, R. & Salinas, E. Sensing and deciding in the somatosensory system. *Curr. Opin. Neurobiol.* **9**, 487–493 (1999).
10. Cohen, M. R. & Newsome, W. T. What electrical microstimulation has revealed about the neural basis of cognition. *Curr. Opin. Neurobiol.* **14**, 169–177 (2004).
11. Stoney, S. D. Jr, Thompson, W. D. & Asanuma, H. Excitation of pyramidal tract cells by intracortical microstimulation: effective extent of stimulating current. *J. Neurophysiol.* **31**, 659–669 (1968).
12. Asanuma, H., Arnold, A. & Zarzecki, P. Further study on the excitation of pyramidal tract cells by intracortical microstimulation. *Exp. Brain Res.* **26**, 443–461 (1976).
13. Tolias, A. S. et al. Mapping cortical activity elicited with electrical microstimulation using fMRI in the macaque. *Neuron* **48**, 901–911 (2005).
14. Nichols, M. J. & Newsome, W. T. Middle temporal visual area microstimulation influences veridical judgments of motion direction. *J. Neurosci.* **22**, 9530–9540 (2002).
15. DeAngelis, G. C. & Newsome, W. T. Perceptual “read-out” of conjoined direction and disparity maps in extrastriate area MT. *PLoS Biol.* **2**, E77 (2004).
16. Perrett, D. I., Hietanen, J. K., Oram, M. W. & Benson, P. J. Organization and functions of cells responsive to faces in the temporal cortex. *Phil. Trans. R. Soc. Lond. B* **335**, 23–30 (1992).
17. Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B. & Tootell, R. B. Faces and objects in macaque cerebral cortex. *Nature Neurosci.* **6**, 989–995 (2003).
18. Zangenehpour, S. & Chaudhuri, A. Patchy organization and asymmetric distribution of the neural correlates of face processing in monkey inferotemporal cortex. *Curr. Biol.* **15**, 993–1005 (2005).
19. Salzman, C. D., Britten, K. H. & Newsome, W. T. Cortical microstimulation influences perceptual judgements of motion direction. *Nature* **346**, 174–177 (1990); Erratum. *Nature* **346**, 589 (1990).
20. Moore, T. & Armstrong, K. M. Selective gating of visual signals by microstimulation of frontal cortex. *Nature* **421**, 370–373 (2003).
21. Bruce, C., Desimone, R. & Gross, C. G. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* **46**, 369–384 (1981).
22. Sugase, Y., Yamane, S., Ueno, S. & Kawano, K. Global and fine information coded by single neurons in the temporal visual cortex. *Nature* **400**, 869–873 (1999).
23. Kiani, R., Esteky, H. & Tanaka, K. Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *J. Neurophysiol.* **94**, 1587–1596 (2005).
24. Horel, J. A., Pytko-Joiner, D. E., Voytko, M. L. & Salisbury, K. The performance of visual tasks while segments of the inferotemporal cortex are suppressed by cold. *Behav. Brain Res.* **23**, 29–42 (1987).
25. Buffalo, E. A. et al. Dissociation between the effects of damage to perirhinal cortex and area TE. *Learn. Mem.* **6**, 572–599 (1999).
26. Tovee, M. J., Rolls, E. T., Treves, A. & Bellis, R. P. Information encoding and the responses of single neurons in the primate temporal visual cortex. *J. Neurophysiol.* **70**, 640–654 (1993).
27. Keyser, C., Xiao, D. K., Foldiak, P. & Perrett, D. I. The speed of sight. *J. Cogn. Neurosci.* **13**, 90–101 (2001).
28. Foldiak, P., Xiao, D., Keyser, C., Edwards, R. & Perrett, D. I. Rapid serial visual presentation for the determination of neural selectivity in area STSa. *Prog. Brain Res.* **144**, 107–116 (2004).
29. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Peninsula Publishing, Los Altos, California, 1988).
30. Meeker, W. Q. & Escobar, L. A. Teaching about approximate confidence regions based on maximum likelihood estimation. *Am. Stat.* **49**, 48–53 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S.-M. Noorbakhsh for technical assistance, and Y. Najian for training one of the monkeys. We also thank W. T. Newsome, M. N. Shadlen and N. Kanwisher for comments on the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.E. (esteky@ipm.ir).

LETTERS

Enzymatic activation of voltage-gated potassium channels

Yajamana Ramu¹, Yanping Xu¹ & Zhe Lu¹

Voltage-gated ion channels in excitable nerve, muscle, and endocrine cells generate electric signals in the form of action potentials¹. However, they are also present in non-excitable eukaryotic cells and prokaryotes, which raises the question of whether voltage-gated channels might be activated by means other than changing the voltage difference between the solutions separated by the plasma membrane. The search for so-called voltage-gated channel activators is motivated in part by the growing importance of such agents in clinical pharmacology. Here we report the apparent activation of voltage-gated K⁺ (Kv) channels by a sphingomyelinase.

While screening various sources we found that the venom of the brown spider, *Loxosceles reclusa*, causes Kv channels (for example, Kv2.1 (ref. 2)) to conduct robust outward current at potentials typical for many resting cells, where they otherwise remain largely deactivated (Fig. 1a). The same effect is observed in a scorpion toxin-sensitive mutant Kv2.1 channel (Kv2.1-Δ7 (ref. 3)) that contains seven mutations in the linker between the fifth and sixth transmembrane segments (S5 and S6) (Fig. 1b), and which we used for the studies described below.

To purify the channel-activating activity, we ran the venom through a size-exclusion fast-flow and low-pressure liquid chromatography (FPLC) column and subjected one of the active fractions to reverse-phase high-performance liquid chromatography (HPLC) (Fig. 1c, d). Analysis by HPLC coupled with tandem mass spectrometry (LC-MS/MS) of a trypsin digest of the HPLC-purified material(s) identified fragments (see Supplementary Fig. 1) that matched those of the Lr1 (11 peptides) and Lr2 (8 peptides) isoforms⁴ of an enzyme commonly called sphingomyelinase D (SMase D), which is known to be present in the venom and to cause severe dermonecrosis^{5,6} (some bacteria also produce SMase D^{7–9}). The observed masses of the active materials (31.15 and 31.36 kDa) matched the theoretical masses of Lr1 and Lr2 (31.18 and 31.33 kDa). We infer that the active agents in the venom are in fact SMase D isoforms. Supporting this conclusion, the recombinant Lr2 isoform¹⁰, like the native materials, induces robust outward current in the wild-type and mutant Kv2.1 channels while the membrane is held at –50 mV throughout (Fig. 1e, f; see Methods for the production of recombinant Lr2). Furthermore, we observe no current activation by SMase D that has been pre-exposed to low pH or has had its two critical histidine residues mutated (Fig. 2a, b), two manipulations known to inactivate SMase D^{10–12}.

Although most phenomena discussed below were originally observed with native materials, the actual data shown were gathered subsequently with recombinant SMase D (–Lr2) (4 ng μl^{–1} unless otherwise specified). SMase D catalyses a reaction in which the positively charged choline group of zwitterionic sphingomyelin is removed, yielding the negatively charged lipid ceramide 1-phosphate⁵ (Fig. 2d; SMase D also catalyses the cleavage of choline from lysophosphatidylcholine bound to albumin in plasma, yielding

lysophosphatidic acid^{7,10,13}). Although phosphatidylinositol-4,5-bisphosphate, which affects the activity of many types of ion channel¹⁴, is naturally present in the inner leaf of the plasma membrane, sphingomyelin is present mainly in the outer leaf. SMase D contains a Mg²⁺ ion in its catalytic centre¹⁵. Like its lipase activity, the channel-activating action of SMase D is Mg²⁺ dependent (Fig. 2c).

The charge on lipid headgroups contributes to the membrane surface potential and affects the functional properties of some K⁺ channels^{16–18}. It is conceivable that the positively charged choline group in the lipid substrates of SMase D, which are located near the outer surface of the membrane, interferes energetically with

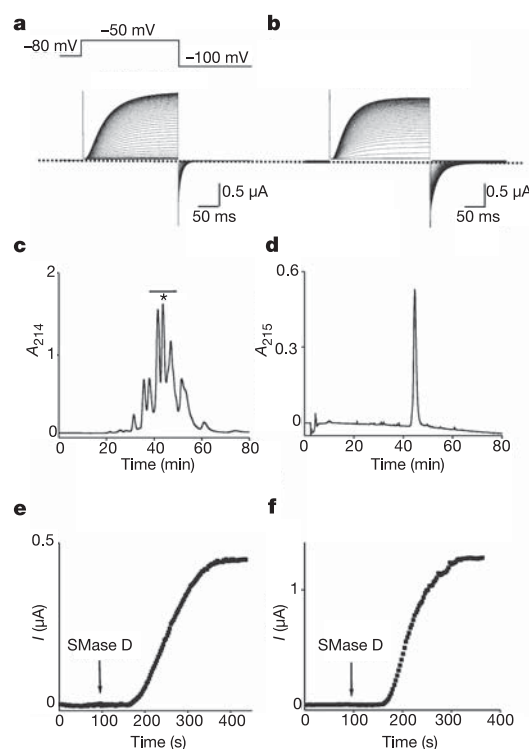


Figure 1 | Identification of Kv channel-stimulating activity. **a, b**, Wild-type (**a**) and Δ7 mutant (**b**) Kv2.1 currents elicited at 3-s intervals with the voltage protocol shown, which gradually appeared after venom addition (1:3,000 dilution). The dotted line indicates the zero current. **c**, Size-exclusion chromatography of 10 μl of venom; the bar identifies active fractions. **d**, The asterisk-marked peak in **c** was further purified on a reverse-phase column, yielding a major peak containing the activity. **e, f**, Kv2.1 (**e**) and Kv2.1-Δ7 (**f**) currents induced by recombinant SMase D (4 ng μl^{–1}) as the membrane voltage V_m was held at –50 mV throughout.

¹Department of Physiology, University of Pennsylvania, 3700 Hamilton Walk, Philadelphia, Pennsylvania 19104, USA.

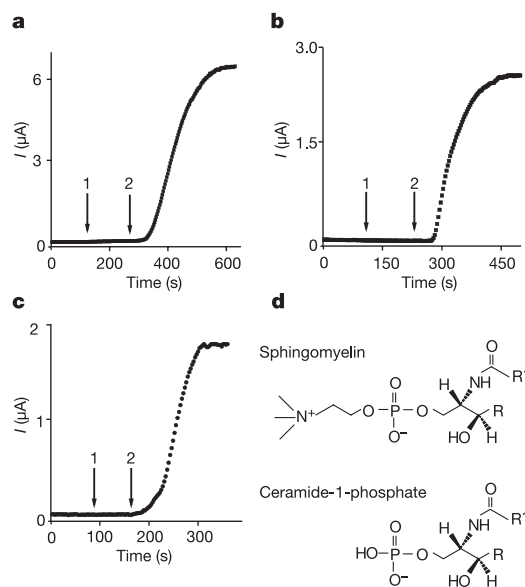


Figure 2 | Effects of pH, histidine mutations and Mg^{2+} on SMase D activity. Amplitudes of Kv2.1-Δ7 currents, repeatedly elicited as in Fig. 1a, are plotted against time. **a**, Arrows indicate successive addition of SMase D-Lr2 that was preincubated (1) or not (2) in 0.1% trifluoroacetic acid (pH < 3) for 14 h. **b**, Successive addition of SMase D-Lr2 with (1) or without (2) the H11A and H47A mutations. **c**, No Mg^{2+} was present in the bath solution during the first ~3 min. Arrows indicate successive additions of SMase D-Lr2 (1) and 1 mM Mg^{2+} (2). **d**, Chemical structures of sphingomyelin and ceramide-1-phosphate; R and R' indicate acyl chains.

depolarization-induced conformational changes and/or with the movement of the voltage sensors, which contain many positively charged residues (reviewed, for example, in ref. 19; for crystal structure see ref. 20). If that is so, SMase D-catalysed removal of the choline group will eliminate this interference, perhaps even favouring movement of the voltage sensors. SMase D does indeed cause a -30 -mV shift in the conductance–voltage (G – V) curve, which allows activation of the channels at a more negative—typically resting—potential (Fig. 3a–c), and which, as expected, does not recover even 1 h after enzyme wash-out (open symbols; Fig. 3c). If the channels themselves are not primarily affected by the SMase D treatment, they should remain inhibitable by toxins that interfere with their gating (for example hanatoxins (HaTx) (refs 21, 22)) or block their pore (for example agitoxin 2 (AgTx2) (ref. 23)). This is indeed what we observe (Fig. 3e, f).

As divalent cations are known to shift the G – V curves of voltage-gated channels²⁴—an effect commonly understood to reflect lowering of the membrane's negative surface potential¹—raising their concentration might be expected to mitigate the SMase D effect. We found that a tenfold increase in extracellular Mg^{2+} concentration markedly decreases the SMase D-induced shift in the G – V curve (Fig. 3c, d; this difference does not reflect a difference in SMase D activity because in both cases the enzymatic treatment was performed with 1 mM Mg^{2+}). Nevertheless, it is doubtful that the observed effects of SMase D on gating reflect a change in the distributed surface potential of the membrane bilayer. Rather, for the reasons given below, we conclude that they reflect a change in headgroup charge of lipid molecules that interact intimately and (relatively) specifically with the channels. First, under the present experimental conditions, SMase D has little effect on the voltage-gated mEAG K^+ channel²⁵, or on the mSlo K^+ channel²⁶ that is gated by both voltage and intracellular Ca^{2+} (Fig. 4a–f). Second, SMase D causes the G – V curves of different classic Kv channels to shift by different amounts. For example, it shifts the G – V curve of Shaker channels (with inactivation removed²⁷) by only about -10 mV

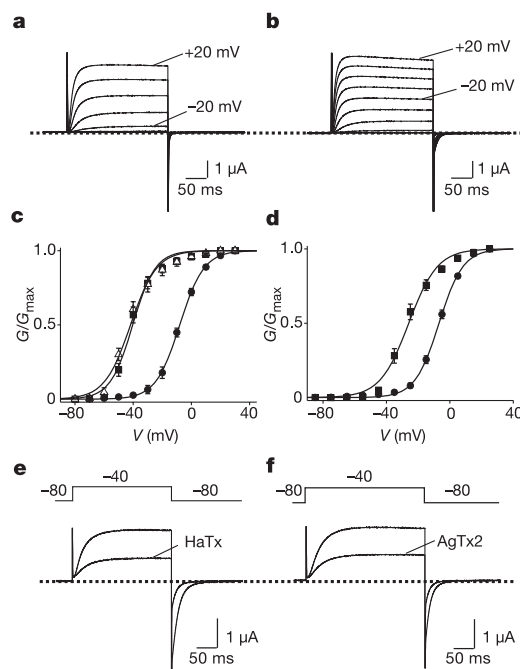


Figure 3 | Effect of SMase D on the G – V curve of Kv2.1-Δ7 channels.

a, b, Currents before (**a**) and after (**b**) SMase D treatment, in which V_m was stepped from the -80 -mV holding voltage (V_{holding}) to various test voltages (V_{test}) up to $+20$ mV in 10 -mV increments ($V_{\text{increment}}$) ($[K^+]_{\text{ext}} = 20$ mM). **c, d**, Plots of G/G_{max} (determined from the tail currents) against V_{test} are fitted with Boltzmann functions, yielding for the case of 1 mM extracellular Mg^{2+} , $V_{1/2} = -8.0 \pm 0.3$ mV (mean $\pm$ s.e.m.; $n = 6$) and $Z = 3.1 \pm 0.1$ before treatment with SMase D (filled circles), $V_{1/2} = -40.0 \pm 0.3$ mV, $Z = 3.3 \pm 0.3$ after treatment with SMase D (filled squares), and $V_{1/2} = -41.8 \pm 0.3$ mV and $Z = 2.9 \pm 0.3$ ($n = 4$) 1 h after SMase D wash-out (open triangles) (**c**), and for the case of 10 mM Mg^{2+} , $V_{1/2} = -7.0 \pm 0.2$ mV ($n = 6$) and $Z = 3.5 \pm 0.2$ before the treatment (filled circles) and $V_{1/2} = -26.0 \pm 0.8$ mV and $Z = 2.9 \pm 0.2$ after the treatment (filled squares) (**d**). **e, f**, Currents (after treatment with SMase D, $[Mg^{2+}]_{\text{ext}} = 1$ mM, $[K^+]_{\text{ext}} = 10$ mM) evoked by stepping V_m from V_{holding} of -80 mV to -40 mV and then back to -80 mV without or with 25 nM HaTx (**e**) and 0.2 nM AgTx2 (**f**).

(Fig. 4g, i). Incidentally but significantly, the non-conducting W434F Shaker channel²⁸ allows one to observe the SMase D-caused increase in gating current at a hyperpolarized potential, as well as a roughly -10 mV shift in the gating charge–voltage (Q – V) curve (Fig. 4h, j; $8 \text{ ng } \mu\text{l}^{-1}$ SMase D was used to reduce the recording time required). Third, the sphingomyelin/phosphatidylcholine abundance ratio in oocyte membranes is only 1:13 (ref. 29) (lysophosphatidylcholine is even scarcer, if present at all). If this ratio holds under our conditions and if phosphatidylcholine is indeed not a substrate for SMase D^{7,10,12,13}, the fact that removing the choline group of sphingomyelin had a significant impact must mean that the association between these lipids and the channel is (relatively) specific. Lipids associated with a K^+ channel protein have actually been resolved in the crystal structure of KcsA (ref. 18), whose transmembrane portion can substitute for the pore module of a Kv channel³⁰. All these findings are consistent with the presence of specific lipid molecules between the voltage-sensing and the pore modules, where oxygen atoms of those lipids may interact with the positively charged residues in the S4. The protein–lipid–protein interaction might explain in part why lipids seem to help crystallized Kv1.2 to maintain a more native conformation²⁰ and indicates that the use of specific lipids might improve the chances of crystallizing voltage-gated channels in desirable gating states.

Thus, SMase D, by cleaving the positively charged choline group from specific types of phospholipid, shifts the G – V (and Q – V) curve

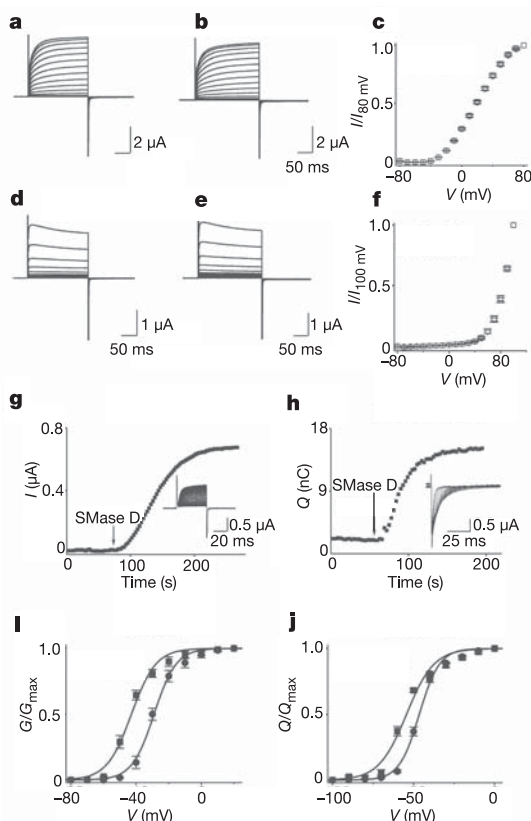


Figure 4 | Effects of SMase D on other K⁺ channels. **a–f**, mEAG (**a**, **b**) and mSlo K⁺ (**d**, **e**) currents elicited by stepping V_m from V_{holding} of -80 mV to various V_{test} values up to $+80$ or $+100$ mV, where $V_{\text{increment}} = 10$ mV before (**a**, **d**) and after (**b**, **e**) treatment with SMase D, and the (normalized) currents (mean $\pm$ s.e.m., $n = 5$) at the various V_{test} values are shown in **c** and **f**, respectively (circles, control; squares, SMase D). **g**, **h**, Time course of current of Shaker(-IR) (elicited by stepping V_m from -80 mV to -50 mV; inset (**g**) and the (off-)gating charges of the W434F mutant (elicited by stepping V_m from -60 mV to -100 mV; inset (**h**)). For gating current studies, the current records were corrected for background ionic and capacitive currents with the P/6 protocol. **i**, **j**, G - V and Q - V curves of Shaker(-IR) (**i**) and W434F mutant (**j**) before (circles) and after (squares) treatment with SMase D ($[K^+]_{\text{ext}} = 100$ mM; $V_{\text{holding}} = -80$ mV, $V_{\text{test}} = -80$ to 20 mV with $V_{\text{increment}} = 10$ mV, $V_{\text{tail}} = -70$ mV for **i**, and $[K^+]_{\text{ext}} = 20$ mM; $V_{\text{holding}} = -100$ mV, $V_{\text{test}} = -100$ to 0 mV with $V_{\text{increment}} = 10$ mV, $V_{\text{tail}} = -100$ mV for **j**) are fitted with Boltzmann functions, yielding $V_{1/2} = -30.0 \pm 0.6$ mV (mean $\pm$ s.e.m.; $n = 8$) and $Z = 3.8 \pm 0.3$ before treatment and $V_{1/2} = -41.2 \pm 0.9$ mV and $Z = 3.5 \pm 0.3$ after treatment (**i**), and $V_{1/2} = -46.5 \pm 0.5$ mV (mean $\pm$ s.e.m.; $n = 7$) and $Z = 4.2 \pm 0.3$ before treatment and $V_{1/2} = -54.5 \pm 1.0$ mV and $Z = 3.6 \pm 0.3$ after treatment (**j**).

of Kv channels in the negative direction, effectively activating the channels at negative potentials at which they would otherwise remain practically deactivated and conduct little current. This previously unknown means of activation of voltage-gated ion channels provides significant insight into how cell excitability may be regulated (or affected pathophysiologically) by altering the properties of lipids in (the outer leaf of) the plasma membrane, and how voltage-gated channels may be activated in non-excitabile cells or lower organisms. Additionally, SMase D offers a simple, effective pharmacological way to activate certain voltage-gated channels at membrane potentials near resting.

METHODS

Molecular biology and electrophysiological recordings. The complementary DNAs of Kv2.1 and mSlo were cloned in pBluescript and pcDNA3 plasmids, respectively, whereas Shaker H4 with N-type inactivation removed (Shaker-IR) and mEAG were cloned in the pGEM-HES plasmid. All mutant cDNAs were

obtained through PCR-based mutagenesis and confirmed by DNA sequencing. The cRNAs were synthesized with T7 or SP6 polymerases, using the corresponding linearized cDNAs as templates. Channel currents were recorded from whole oocytes (previously injected with the corresponding cRNA) with a two-electrode voltage-clamp amplifier (Warner OC-725C). The resistance of electrodes filled with 3 M KCl was 0.2 – 0.3 M Ω . Unless specified otherwise, the bath solution contained (in mM): 95 Na⁺ ($\text{Cl}^- + \text{OH}^-$), 5 KCl, 0.3 CaCl₂, 1 MgCl₂ and 10 HEPES; pH was adjusted to 7.6 with NaOH. When K⁺ concentration was increased, Na⁺ concentration was decreased to keep their sum at 100 mM. SMase D (1 – 2 μ l) was added manually to the recording chamber (100 μ l).

Identification, purification and production of recombinant SMase D. For purification, *Loxosceles reclusa* venom samples were loaded on a size-exclusion FPLC column (Superdex G200; Pharmacia) in which the running buffer contained 50 mM NaCl and 5 mM Tris-HCl (pH 6.4). One of the active fractions was subsequently loaded on a reverse-phase C₁₈ HPLC column (Beckman) and eluted with a water and acetonitrile gradient (increasing by 1% acetonitrile per minute). The aqueous and organic mobile phases contained 0.1% and 0.07% trifluoroacetic acid, respectively. The sample corresponding to the main peak on the HPLC chromatogram contained the activity and was subsequently analysed by matrix-assisted laser desorption/ionization–time-of-flight MS for mass identification. The trypsin digestion products of the HPLC-purified material were then analysed with LC–MS/MS, yielding 11 and 8 partial peptide sequences corresponding to the Lr1 and Lr2 isoforms of SMase D (Supplementary Fig. 1). The SMase D activity was drastically decreased after HPLC purification because of exposure to low pH^{11,12}. To produce the recombinant mature form of wild-type and the H11A + H47A mutant SMase D, *Escherichia coli* BL21 (DE3) cells were transformed with the respective cDNA of the Lr2 isoform cloned in pET30 Ek/LIC vector, grown in Luria–Bertani medium to a D_{600} of about 0.6 , and induced for 2 h with 1 mM isopropyl β -D-thiogalactoside. The bacteria were harvested, resuspended and sonicated. The resulting sample was loaded onto a cobalt-affinity column and eluted by stepping the imidazole concentration from 50 mM to 500 mM. The imidazole was later removed by dialysis.

Received 28 February; accepted 8 May 2006.

Published online 21 June 2006.

- Hille, B. *Ion Channels of Excitable Membranes* 3rd edn (Sinauer, Sunderland, Massachusetts, 2001).
- Frech, G. C., VanDongen, A. M., Schuster, G., Brown, A. M. & Joho, R. H. A novel potassium channel with delayed rectifier properties isolated from rat brain by expression cloning. *Nature* **340**, 642–645 (1989).
- Aggarwal, S. K. *Analysis of the Voltage-sensor in a Voltage-activated Potassium Channel*. PhD thesis, Harvard Univ. (1996).
- Ramos-Cerrillo, B. *et al.* Genetic and enzymatic characterization of sphingomyelinase D isoforms from the North American fiddleback spiders *Loxosceles boneti* and *Loxosceles reclusa*. *Toxicon* **44**, 507–514 (2004).
- Kurpiewski, G., Forrester, L. J., Barrett, J. T. & Campbell, B. J. Platelet aggregation and sphingomyelinase D activity of a purified toxin from the venom of *Loxosceles reclusa*. *Biochim. Biophys. Acta* **678**, 467–476 (1981).
- Rees, R. S., Nanney, L. B., Yates, R. A. & King, L. E. Jr. Interaction of brown recluse spider venom on cell membranes: the inciting mechanism? *J. Invest. Dermatol.* **83**, 270–275 (1984).
- Soucek, A., Michalec, C. & Souckova, A. Identification and characterization of a new enzyme of the group 'phospholipase D' isolated from *Corynebacterium ovis*. *Biochim. Biophys. Acta* **227**, 116–128 (1971).
- Barksdale, L., Linder, R., Sulea, I. T. & Pollice, M. Phospholipase D activity of *Corynebacterium pseudotuberculosis* (*Corynebacterium ovis*) and *Corynebacterium ulcerans*, a distinctive marker within the genus *Corynebacterium*. *J. Clin. Microbiol.* **13**, 335–343 (1981).
- Bernheimer, A. W., Campbell, B. J. & Forrester, L. J. Comparative toxinology of *Loxosceles reclusa* and *Corynebacterium pseudotuberculosis*. *Science* **228**, 590–591 (1985).
- Lee, S. & Lynch, K. R. Brown recluse spider (*Loxosceles reclusa*) venom phospholipase D (PLD) generates lysophosphatidic acid (LPA). *Biochem. J.* **391**, 317–323 (2005).
- de Andrade, S. A. *et al.* Conformational changes of *Loxosceles* venom sphingomyelinases monitored by circular dichroism. *Biochem. Biophys. Res. Commun.* **327**, 117–123 (2005).
- de Andrade, S. A., Murakami, M. T., Cavalcante, D. P., Arni, R. K. & Tambourgi, D. V. Kinetic and mechanistic characterization of the sphingomyelinases D from *Loxosceles intermedia* spider venom. *Toxicon* **47**, 380–386 (2006).
- van Meeteren, L. A. *et al.* Spider and bacterial sphingomyelinases D target cellular lysophosphatidic acid receptors by hydrolyzing lysophosphatidylcholine. *J. Biol. Chem.* **279**, 10833–10836 (2004).
- Hilgemann, D. W., Feng, S. & Nasuhoglu, C. The complex and intriguing lives of PIP₂ with ion channels and transporters. *Sci. STKE* **2001**, RE19 (2001).
- Murakami, M. T., Fernandes-Pedrosa, M. F., Tambourgi, D. V. & Arni, R. K. Structural basis for metal ion coordination and the catalytic mechanism of sphingomyelinases D. *J. Biol. Chem.* **280**, 13658–13664 (2005).

16. Bell, J. E. & Miller, C. Effects of phospholipid surface charge on ion conduction in the K⁺ channel of sarcoplasmic reticulum. *Biophys. J.* **45**, 279–287 (1984).
17. Moczydlowski, E., Alvarez, O., Vergara, C. & Latorre, R. Effect of phospholipid surface charge on the conductance and gating of a Ca²⁺-activated K⁺ channel in planar lipid bilayers. *J. Membr. Biol.* **83**, 273–282 (1985).
18. Valiyaveetil, F. I., Zhou, Y. & MacKinnon, R. Lipids in the structure, folding, and function of the KcsA K⁺ channel. *Biochemistry* **41**, 10771–10777 (2002).
19. Bezanilla, F. The voltage sensor in voltage-dependent ion channels. *Physiol. Rev.* **80**, 555–592 (2000).
20. Long, S. B., Campbell, E. B. & MacKinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel. *Science* **309**, 897–903 (2005).
21. Swartz, K. J. & MacKinnon, R. An inhibitor of the Kv2.1 potassium channel isolated from the venom of a Chilean tarantula. *Neuron* **15**, 941–949 (1995).
22. Phillips, L. R. *et al.* Voltage-sensor activation with a tarantula toxin as cargo. *Nature* **436**, 857–860 (2005).
23. Garcia, M. L., Garcia-Calvo, M., Hidalgo, P., Lee, A. & MacKinnon, R. Purification and characterization of three inhibitors of voltage-dependent K⁺ channels from *Leiurus quinquestriatus* var. *hebraeus* venom. *Biochemistry* **33**, 6834–6839 (1994).
24. Frankenhaeuser, B. & Hodgkin, A. L. The action of calcium on the electrical properties of squid axons. *J. Physiol. (Lond.)* **137**, 218–244 (1957).
25. Robertson, G. A., Warmke, J. M. & Ganetzky, B. Potassium currents expressed from *Drosophila* and mouse eag cDNAs in *Xenopus* oocytes. *Neuropharmacology* **35**, 841–850 (1996).
26. Butler, A., Tsunoda, S., McCobb, D. P., Wei, A. & Salkoff, L. mSlo, a complex mouse gene encoding 'maxi' calcium-activated potassium channels. *Science* **261**, 221–224 (1993).
27. Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. *Science* **250**, 533–538 (1990).
28. Perozo, E., MacKinnon, R., Bezanilla, F. & Stefani, E. Gating currents from a nonconducting mutant reveal open-closed conformations in Shaker K⁺ channels. *Neuron* **11**, 353–358 (1993).
29. Stith, B. J. *et al.* Quantification of major classes of *Xenopus* phospholipids by high performance liquid chromatography with evaporative light scattering detection. *J. Lipid Res.* **41**, 1448–1454 (2000).
30. Lu, Z., Klem, A. M. & Ramu, Y. Ion conduction pore is conserved among potassium channels. *Nature* **413**, 809–813 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Lynch (University of Virginia) for sharing the cDNA clone of Lr2 isoform of SMase D; R. Joho and K. Swartz for Kv2.1 cDNA; G. Robertson and B. Ganetzky for mEAG cDNA; L. Salkoff and F. Horrigan for mSlo cDNA; K. Swartz for Shaker-IR cDNA in the pGEM-HES vector and the HaTx sample; C.-X. Yuan for LC-MS/MS sequencing; J. R. Martinez-Francois for chemical structure drawing; R. MacKinnon for comments; and P. De Weer for review and discussion of our manuscript. This study was supported by a grant from the National Institute of General Medical Sciences to Z.L.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Z.L. (zhelu@mail.med.upenn.edu).

LETTERS

An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis

Fajun Yang^{1,2*}, Bryan W. Vought^{3*}, John S. Satterlee¹, Amy K. Walker^{1,4}, Z.-Y. Jim Sun³, Jennifer L. Watts⁵, Rosalie DeBeaumont^{1,2}, R. Mako Saito^{1,4†}, Sven G. Hyberts³, Shaosong Yang^{1,2†}, Christine Macol^{1,2†}, Lakshmanan Iyer⁶, Robert Tjian⁷, Sander van den Heuvel^{1,4†}, Anne C. Hart^{1,4}, Gerhard Wagner³ & Anders M. Näär^{1,2}

The sterol regulatory element binding protein (SREBP) family of transcription activators are critical regulators of cholesterol and fatty acid homeostasis^{1,2}. We previously demonstrated that human SREBPs bind the CREB-binding protein (CBP)/p300 acetyltransferase KIX domain and recruit activator-recruited co-factor (ARC)/Mediator co-activator complexes through unknown mechanisms^{3–5}. Here we show that SREBPs use the evolutionarily conserved ARC105 (also called MED15) subunit to activate target genes. Structural analysis of the SREBP-binding domain in ARC105 by NMR revealed a three-helix bundle with marked similarity to the CBP/p300 KIX domain. In contrast to SREBPs, the CREB and c-Myb activators do not bind the ARC105 KIX domain, although they interact with the CBP KIX domain, revealing a surprising specificity among structurally related activator-binding domains. The *Caenorhabditis elegans* SREBP homologue SBP-1 promotes fatty acid homeostasis by regulating the expression of lipogenic enzymes^{6,7}. We found that, like SBP-1, the *C. elegans* ARC105 homologue MDT-15 is required for fatty acid homeostasis, and show that both SBP-1 and MDT-15 control transcription of genes governing desaturation of stearic acid to oleic acid. Notably, dietary addition of oleic acid significantly rescued various defects of nematodes targeted with RNA interference against *sbp-1* and *mdt-15*, including impaired intestinal fat storage, infertility, decreased size and slow locomotion, suggesting that regulation of oleic acid levels represents a physiologically critical function of SBP-1 and MDT-15. Taken together, our findings demonstrate that ARC105 is a key effector of SREBP-dependent gene regulation and control of lipid homeostasis in metazoans.

Cholesterol and fatty acids have important functional roles in metazoans, such as modulating membrane fluidity, serving as signaling molecules and providing energy storage in the form of triacylglycerides. Abnormal cholesterol and fat levels have been linked to prevalent diseases, including atherosclerosis, obesity, type 2 diabetes and hypertension (all associated with metabolic syndrome), underscoring the importance of understanding fully how cholesterol and lipid homeostasis are regulated and maintained^{8–10}.

The mammalian SREBP family of basic helix–loop–helix zipper (bHLH–Zip) transcription factors promote adipocyte differentiation and are critical regulators of cholesterol and fatty acid homeostasis

by controlling the expression of cholesterologenic and lipogenic genes^{1,2}. Human SREBPs can activate target genes by recruiting the chromatin-targeting CBP/p300 acetyltransferases and the polymerase-II-interacting ARC/Mediator co-activators^{3–5,11}. A short sequence within the amino-terminal activation domain of SREBP-1a bound both CBP/p300 and the ARC/Mediator co-activator (Fig. 1a, e). We previously identified the KIX domain in CBP/p300 as a functionally important target of the SREBP activation domain³ (Fig. 1b). Bioinformatics analysis revealed that the ARC105 subunit of the ARC/Mediator co-activator contains a region with significant sequence similarity to the CBP/p300 KIX domain¹², raising the possibility that structurally related motifs present in both CBP/p300 and the ARC/Mediator co-activator might mediate interaction with SREBP-1a (Fig. 1b; see also Supplementary Fig. S1). Indeed, ARC105 bound to the SREBP-1a activation domain, but not to a number of other activators^{13,14} (Fig. 1c, e and data not shown). The SREBP-1a activation domain did not bind significantly to other ARC/Mediator subunits tested (Fig. 1c and data not shown). The other domains of the SREBP-1c and -2 isoforms also interacted with the ARC/Mediator co-activator and ARC105, albeit more weakly than SREBP-1a (Fig. 1e). SREBP-1a associated with ARC105 *in vivo*, as revealed by co-immunoprecipitation studies (Fig. 1d). Together, these results show that ARC105 is a direct target of the SREBP family of activators.

The N-terminal region of ARC105 containing the predicted KIX domain strongly interacted with recombinant SREBP-1a (Fig. 2c), supporting the hypothesis that SREBPs associate with distinct co-activators that harbour related KIX domains. Fluorescence polarization analysis also showed that the SREBP-1a activation domain interacts with high affinity (dissociation constant (K_d) of ~120 nM) with the putative ARC105 KIX domain (Supplementary Fig. S2). Structural studies with NMR clearly demonstrated that the ARC105 SREBP-binding domain folds into a three-helix bundle with striking similarity to the CBP KIX domain (Fig. 2a, b). Deletion analysis and point mutagenesis provided evidence that the third α -helix of the ARC105 KIX domain contains the primary SREBP-1a-binding site (Fig. 2c, e, f and data not shown).

The CBP KIX domain mediates interactions with a number of activators, including the cAMP/protein kinase A (PKA)-regulated CREB and the c-Myb transactivators^{15,16}. However, although the

¹Massachusetts General Hospital Cancer Center, Building 149, 13th Street, Charlestown, Massachusetts 02129, USA. ²Department of Cell Biology, ³Department of Biological Chemistry and Molecular Pharmacology, and ⁴Department of Pathology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. ⁵Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164, USA. ⁶Bauer Center for Genomics Research, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA. ⁷Department of Molecular and Cell Biology, UC Berkeley, Berkeley, California 94720, USA. [†]Present addresses: Department of Genetics, Dartmouth Medical School, 7400 Remsen, Hanover, New Hampshire 03755, USA (R.M.S.); ERDC-CERL, PO Box 9005, Champaign, Illinois 61826-9005, USA (S.Y.); Prince George's Community College, 301 Largo Road, Largo, Maryland 20774, USA (C.M.); Utrecht University, Department of Developmental Biology, Kruytgebouw N305, Padualaan 8, 3584 CH, Utrecht, The Netherlands (S.v.d.H.).

*These authors contributed equally to this work.

activation domains of CREB and c-Myb bound strongly to the CBP KIX domain, they did not exhibit significant binding to the ARC/Mediator co-activator, nor to the ARC105 subunit or its KIX domain (Figs 1a and 2d and data not shown). Thus, the ARC105 KIX domain interacts with only a subset of CBP/p300 KIX-binding activators. Tyr 658 and Lys 662 in the CBP KIX domain have been implicated in the interaction with PKA-phosphorylated CREB and c-Myb activation domains^{17–20}; however, these amino acids are not conserved in the ARC105 KIX domain, potentially explaining why CREB and the c-Myb activation domain did not bind the ARC105 KIX domain (Fig. 2e, f). To address this possibility, Ile64 and Asp68 in the ARC105 KIX domain were changed into the corresponding CBP amino acids (I64Y/D68K). The doubly mutated ARC105 KIX domain interacted with both CREB and the activation domain of c-Myb fused to the Gal4 DNA-binding domain (Gal4-MybAD), albeit more weakly than with the CBP KIX domain (Fig. 2d, e).

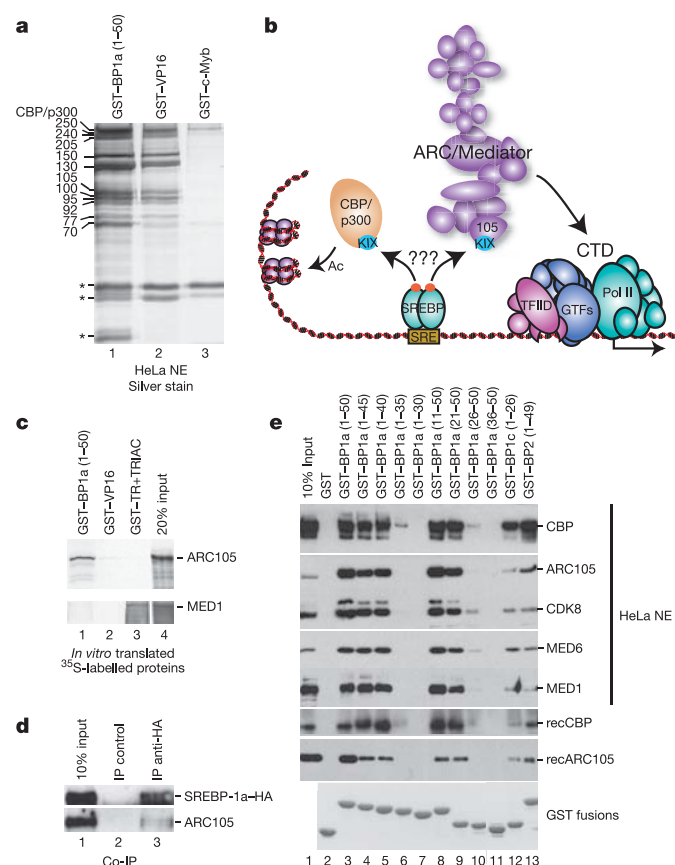


Figure 1 | The activation domains of SREBPs interact with the ARC105 subunit of the ARC/Mediator co-activator. **a**, The ARC/Mediator co-activator from HeLa cell nuclear extract (NE) bound to GST fusions of the activation domains of human SREBP-1a (BP1a) and VP16, but not to human c-Myb. All three activation domains interacted with the CBP/p300 co-activators. Molecular weights (kDa) of ARC/Mediator subunits are indicated to the left of the panel. Non-specific proteins are indicated with asterisks. **b**, Cartoon illustrating the question of whether the small SREBP activation domain recruits both CBP/p300 and the ARC/Mediator via KIX domains. **c**, The SREBP-1a activation domain bound to the ARC105 subunit, whereas the ligand-bound thyroid hormone receptor (TR) specifically interacted with MED1 (ref. 30). **d**, Haemagglutinin-tagged SREBP-1a (SREBP-1a-HA) expressed in U2OS cells co-immunoprecipitated with endogenous ARC105. **e**, CBP and the intact ARC/Mediator co-activator from nuclear extract, as well as recombinant CBP and ARC105, interact with the same sequence within the SREBP-1a activation domain. The activation domains of SREBP-1c and SREBP-2 also interacted with CBP and ARC105.

These results suggest that several amino acids located primarily in the third α -helix of the CBP and ARC105 KIX domains provide activator-binding specificity. Notably, chemical shift analysis revealed that several amino acid resonances in the third helix were perturbed upon binding unlabelled SREBP, indicating that this area of the ARC105 KIX domain is the key SREBP docking site (Fig. 2e, f). The SREBP-1a binding site on the ARC105 KIX domain appears to be distinct from the surfaces on CBP KIX that are bound by CREB (pKID), c-Myb and in particular, MLL (Fig. 2f)^{17–22}. Together with the lack of conservation of Tyr 658 and Lys 662 in the ARC105 KIX domain, these differences may account for why CREB (pKID) and c-Myb are unable to bind effectively to the ARC105 KIX domain (Fig. 2f).

Consistent with the findings suggesting that the third α -helix of the ARC105 KIX domain constitutes the primary SREBP-binding site, addition of a peptide containing the third α -helix strongly inhibited interaction of the SREBP-1a activation domain with both the ARC/Mediator co-activator and CBP/p300 (Fig. 3a). Moreover, the ARC105 peptide acted as a potent inhibitor of ARC/Mediator-dependent transactivation by SREBP-1a and Sp1 in a chromatin-based *in vitro* transcription reaction, while having no effect on transactivation by Sp1 alone (Fig. 3b; see also Supplementary Fig. S3). Together, these results show that ARC105 harbours a functionally important SREBP-binding domain that is structurally highly similar to the activator-targeted KIX domain found in the CBP/p300 acetyltransferases.

RNA interference (RNAi) was used to investigate the functional role of ARC105 in SREBP-dependent gene activation in human cells. Depletion of ARC105 by short interfering RNA (siRNA) strongly decreased cholesterol-regulated transcription of SREBP target genes, including low-density lipoprotein receptor (*LDLR*), fatty acid synthase (*FASN*), HMG-CoA synthase (*HMG-CoAS*) and HMG-CoA reductase (*HMG-CoAR*)^{1,2} (Fig. 3c), as well as activation of a *FASN* promoter reporter by co-transfected SREBP-1a (Fig. 3d). In contrast, RNAi of *ARC105* did not affect gene activation by several other ARC/Mediator-targeting activators, including BMP-regulated SMAD1 (ref. 13), the viral activator VP16 (ref. 14) and the tumour suppressor p53 (data not shown), indicating that depletion of ARC105 does not globally impinge on the structure or function of the ARC/Mediator co-activator. Consistent with the lack of binding of c-Myb to ARC105, knockdown of ARC105 had no effect on transactivation by the activation domain of c-Myb (Fig. 3d). Next, we demonstrated by chromatin immunoprecipitation (ChIP) that epitope-tagged SREBP-1a and ARC105 co-expressed in U2OS cells co-occupy the *LDLR* and *FASN* promoters (Fig. 3e). Moreover, both CBP and the ARC/Mediator co-activator were found to be recruited to the endogenous *LDLR* and *HMG-CoAR* promoters in a SREBP-1a-dependent manner (Supplementary Fig. S4). These studies show that ARC105 co-localizes with SREBP on target genes and is specifically required for cholesterol-regulated and SREBP-1a-dependent gene activation in human cells.

The ARC/Mediator co-activator is conserved in the nematode *C. elegans*²³, and exploiting the genetic tractability of *C. elegans*, we have investigated the *in vivo* role of the ARC105 homologue MDT-15 in gene activation by the SREBP homologue SBP-1 and in the control of lipid homeostasis^{6,7,24}. The activation domain of SBP-1 stimulated transcription in human cells and interacted with the conserved KIX domains of both MDT-15 and human ARC105, suggesting that the regulatory circuitry observed in human cells may be conserved in *C. elegans* (Fig. 4a; see also Supplementary Fig. S5). RNAi of either *mdt-15* or *sbp-1* produced a transparent or 'clear' phenotype, resembling starved *C. elegans* deprived of nutrients^{7,25} (Fig. 4f). This phenotype has been suggested to possibly correlate with decreased fat storage in the intestinal compartment⁷. By contrast, RNAi depletion of many other ARC/Mediator subunits did not yield a clear intestinal phenotype, indicating that *mdt-15* may serve a specific role in maintaining lipid homeostasis and promoting fat storage (Fig. 4f; see also Supplementary Table 2).

Gas chromatography/mass spectrometry and thin liquid chromatography were used to gain a more detailed understanding of the alterations in lipid homeostasis upon removal of *mdt-15* and *sbp-1*. RNAi of either *mdt-15* or *sbp-1* resulted in decreased levels of triacylglycerides, the main form of fat stored in the intestinal compartment (triacylglycerides as percentage of total lipids: *mdt-15*(RNAi) 32%, *sbp-1*(RNAi) 23%, *mdt-4*(RNAi) 43%, *tbg-1*(RNAi) 47%, empty vector 44%). Moreover, RNAi of *mdt-15* or *sbp-1* brought about similar qualitative changes in the ratios of many different fatty acids, whereas control RNAi had little effect (Fig. 4b; see also Supplementary Fig. S6a).

The saturated fatty acid stearic acid (C18:0) was markedly increased in response to both *sbp-1* and *mdt-15* RNAi, whereas the monounsaturated fatty acid oleic acid was strongly decreased (Fig. 4b). In mammals, SREBPs directly activate transcription of a family of stearoyl-CoA desaturase genes that act in part to desaturate stearic acid to generate oleic acid²⁶. As RNAi of *sbp-1* and *mdt-15* produced a marked increase in the stearic acid to oleic acid ratio (from ~2:1 for vector control to ~7:1 for *sbp-1*(RNAi) and ~16:1 for *mdt-15*(RNAi)), we investigated whether regulation of the *fat-6* and *fat-7* genes, homologues of the mammalian stearoyl-CoA desaturases²⁷, was affected (Supplementary Fig. S6b). Indeed, expression of both *fat-6* and *fat-7* was markedly decreased upon RNAi of *mdt-15* or *sbp-1*, whereas depletion of *C. elegans mdt-4* (control) had little effect, indicating a pivotal and specific role of MDT-15 in the expression of SBP-1-regulated genes governing fatty acid desaturation (Fig. 4c). Moreover, intestinal expression of a *fat-7::GFP* reporter

construct was strongly and specifically reduced upon RNAi of either *sbp-1* or *mdt-15* (Fig. 4d; see also Supplementary Fig. S7).

RNAi of *mdt-15* and *sbp-1* resulted in similar overt phenotypes, including clear intestines, growth defects, infertility, aberrant locomotion and decreased lifespan²⁵ (Fig. 4e, f; see also Supplementary Fig. S9, and data not shown). Oleic acid serves as a precursor for several physiologically important pathways, including synthesis of triacylglycerides destined for fat storage, membrane phospholipids, and polyunsaturated fatty acids (PUFAs)²⁸ (Supplementary Fig. S6b). Thus, the decreased levels of oleic acid due to impaired expression of the *fat-6* and *fat-7* stearoyl-CoA desaturases might contribute to some of the phenotypes in *mdt-15*(RNAi) and *sbp-1*(RNAi) nematodes. Notably, RNAi of both *fat-6* and *fat-7* (*fat-6/fat-7*(RNAi) nematodes) caused phenotypes that were similar to those resulting from RNAi of *mdt-15* and *sbp-1*, including clear intestines, infertility, decreased motility, smaller size and decreased lifespan²⁵ (Fig. 4f and data not shown). Moreover, dietary supplementation with oleic acid improved the fertility, darkened the intestines, increased locomotion and partially rescued the size defect of *mdt-15* and *sbp-1* RNAi-treated nematodes, and rescued all phenotypes of *fat-6/fat-7*(RNAi) animals (Fig. 4e, f; see also Supplementary Fig. S9). By contrast, stearic acid, which is elevated in response to RNAi of *mdt-15* and *sbp-1*, had little effect. These results collectively suggest that oleic acid deprivation due to decreased *fat-6* and *fat-7* desaturase expression is a primary cause of the clear intestines, small size, decreased fecundity and slow locomotion of *mdt-15*(RNAi) and *sbp-1*(RNAi) nematodes.

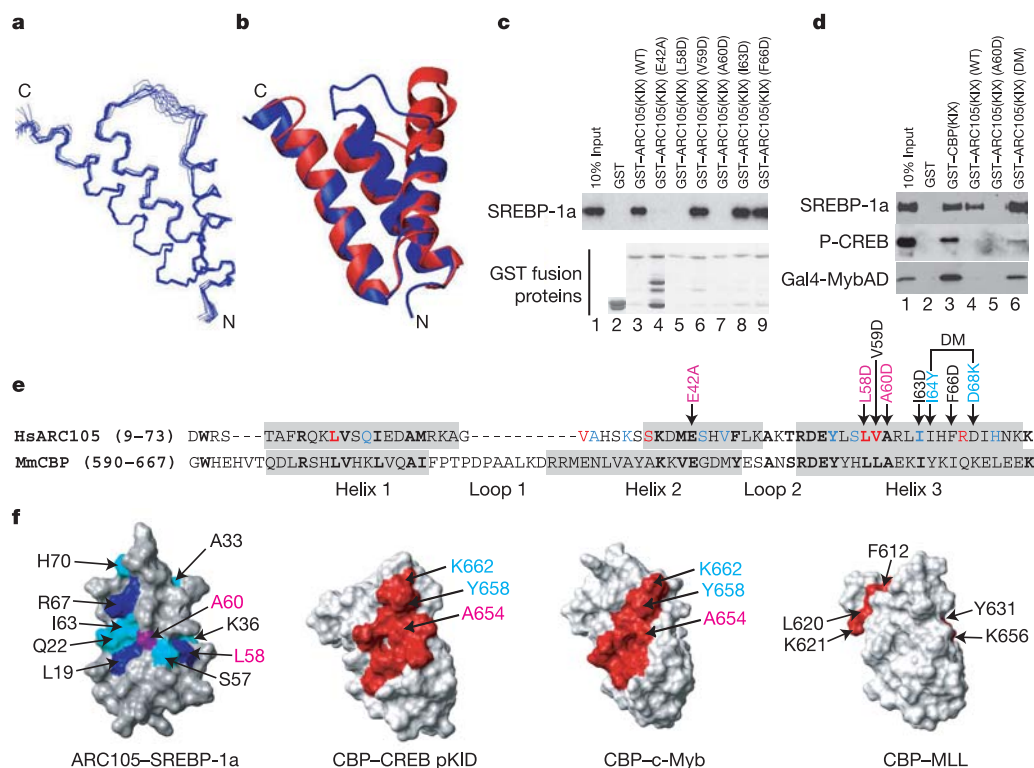


Figure 2 | NMR solution structure of the ARC105 KIX domain.

a, Representation of the 15 lowest energy structures of the ARC105 SREBP-binding domain reveals a three-helix bundle. **b**, The human ARC105 structure (blue) exhibits marked similarity to the mouse CBP KIX domain (red). **c**, Point mutagenesis revealed that E42, L58 and A60 are important for binding of recombinant Flag-tagged SREBP-1a to the ARC105 KIX domain. **d**, PKA-phosphorylated CREB (P-CREB) and Gal4-MybAD bound to immobilized CBP KIX, but not to the ARC105 KIX domain. Flag-SREBP-1a interacted with both KIX domains. Mutating two amino acids in the ARC105 KIX domain to the CBP KIX equivalents (I64Y/D68K, double mutant or DM) allowed partial binding of both P-CREB and Gal4-MybAD.

e, Alignment of the human ARC105 and mouse CBP KIX domains reveals conservation (bold) in the helical domains (grey). Point mutations in the ARC105 KIX domain tested for binding to SREBP-1a and other activators are indicated by arrows. Amino acids coloured red and blue were strongly and moderately affected, respectively, in chemical shift studies upon binding of the SREBP-1a activation domain. **f**, The SREBP-1a binding surface of the ARC105 KIX domain (dark blue/cyan) is distinct from the CBP KIX binding surfaces identified for the activation domains of CREB (pKID), c-Myb and MLL (red). Dark blue amino acids correspond to chemical shift >1.0 and cyan amino acids indicate chemical shift of 0.75–1.0.

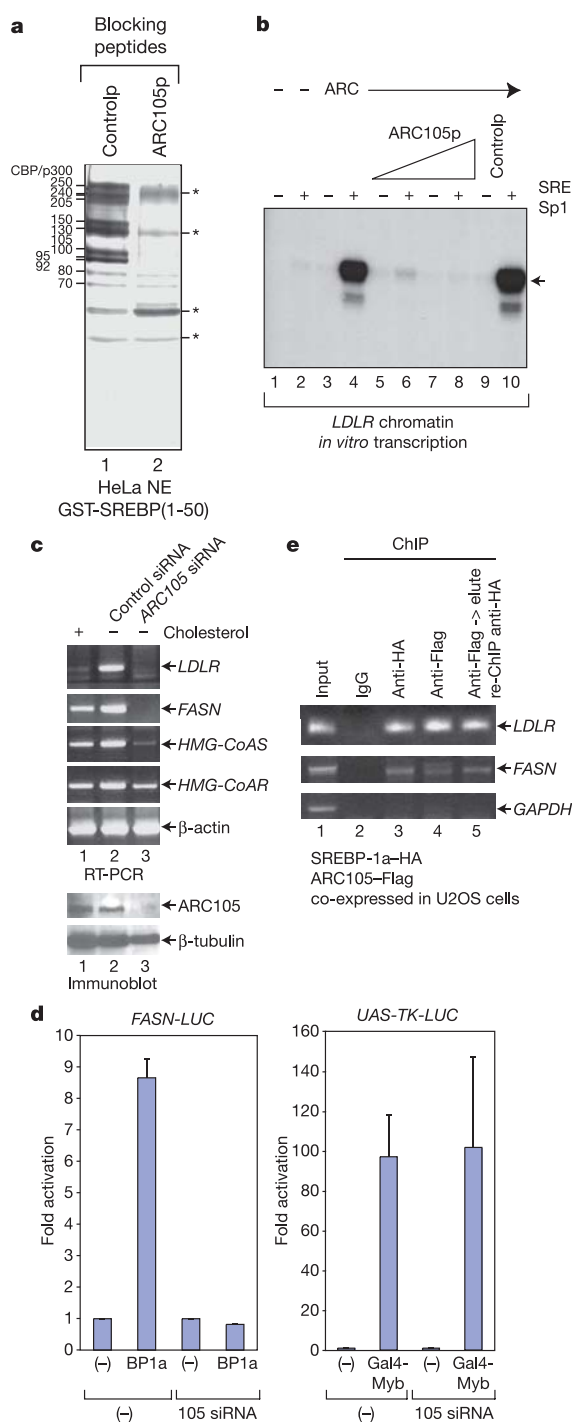


Figure 3 | ARC105 is required for SREBP-dependent transactivation and associates with SREBP target genes in human cells. **a**, An ARC105 KIX peptide (ARC105p) inhibits binding of both CBP/p300 and the ARC/Mediator co-activator to the GST-SREBP-1a activation domain, whereas a control peptide (Controlp) has no effect. Molecular weights and asterisks are as in Fig. 1a. **b**, ARC105p inhibited ARC/Mediator-dependent activation by SREBP-1a and Sp1 in an *in vitro* transcription reaction with *LDLR* promoter-derived chromatin template. **c**, ARC105 knockdown by siRNA in HeLa cells results in decreased cholesterol-regulated expression of SREBP target genes, including *LDLR*, *FASN*, *HMG-CoAS* and *HMG-CoAR*. **d**, ARC105 RNAi inhibited SREBP-1a-dependent transactivation of a *FASN* promoter reporter, without affecting Gal4-MybAD transactivation. **e**, Chromatin immunoprecipitation (ChIP) of SREBP-1a-HA and ARC105-Flag on SREBP target genes in U2OS cells. GAPDH was used as control. Re-ChIP with anti-HA after anti-Flag immunoprecipitation and Flag peptide elution showed that SREBP-1a and ARC105 co-occupy SREBP target genes. Error bars represent s.e.m.

Together with the role of ARC105 in activin/TGF- β -induced gene activation by the SMAD2/3 activators¹³, the findings presented here suggest that ARC105 is a critical transducer of gene activation signals that control early metazoan development as well as cholesterol and fatty acid homeostasis. Tissue-specific changes in the expression of SREBPs, as well as SREBP polymorphisms, have been linked to the development of obesity, atherosclerosis, type 2 diabetes, fatty liver and lipodystrophy in humans and in mice². As ARC105 is a downstream effector of SREBP gene regulation, altered expression levels or mutations/polymorphisms in ARC105 may also contribute to these diseases.

Note added in proof: A recent paper reported a requirement for *mdt-15* in *C. elegans* lipid homeostasis²⁹.

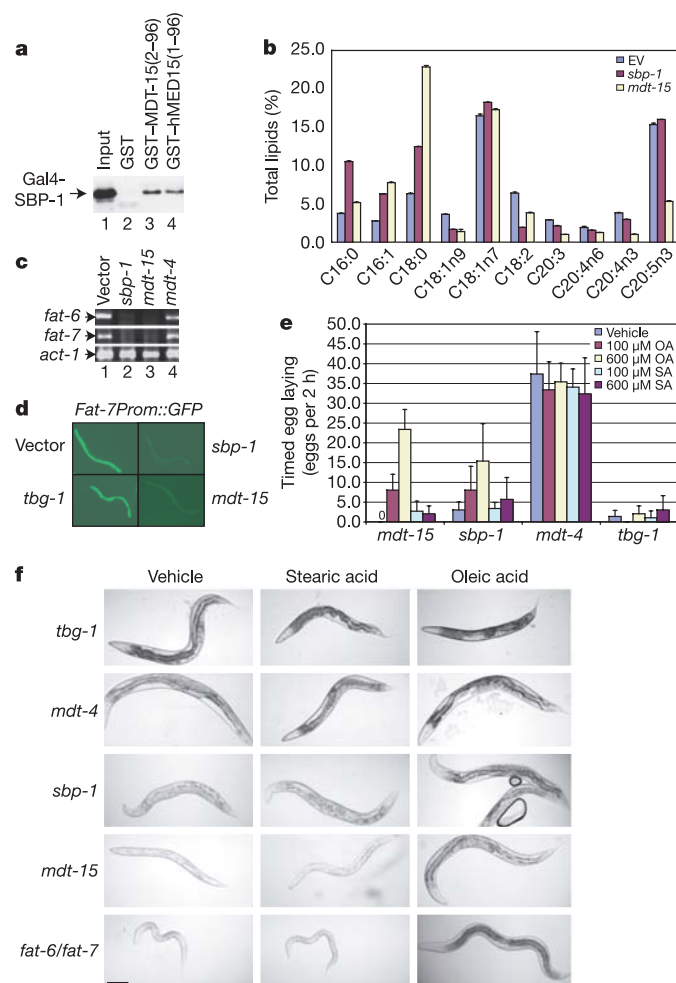


Figure 4 | The *C. elegans* ARC105 homologue MDT-15 is a critical regulator of *C. elegans* SREBP (SBP-1) target genes and is required for normal lipid homeostasis. **a**, The SBP-1 activation domain interacted with the KIX domain of *C. elegans* MDT-15 and human ARC105. **b**, Gas chromatography/mass spectrometry analysis revealed altered fatty acid profiles in *sbp-1(RNAi)* and *mdt-15(RNAi)* nematodes, as compared with empty vector control. **c**, Transcription of the *fat-6* and *fat-7* stearyl-CoA desaturases was strongly decreased upon RNAi of *sbp-1* and *mdt-15*. **d**, Intestinal expression of the *fat-7* promoter fused to GFP (*fat-7Prom::GFP*) decreased after RNAi of *sbp-1* and *mdt-15*. **e**, Dietary addition of oleic acid (OA), but not stearic acid (SA), partially rescued sterility due to *mdt-15* and *sbp-1* RNAi, while having no effect on controls (*tbp-1(RNAi)* and *mdt-4(RNAi)*). **f**, RNAi of *sbp-1*, *mdt-15* and *fat-6* and *fat-7* yielded pale intestines and smaller animals. Dietary addition of 600 μ M oleic acid, but not stearic acid, darkened the intestines and increased the size of *mdt-15(RNAi)*, *sbp-1(RNAi)* and *fat-6/fat-7(RNAi)* nematodes. The scale bar in the lower left panel represents 0.1 mm. Error bars represent s.e.m.

METHODS

Details of plasmids, primers for RT-PCR and ChIP, siRNA, antibodies, expression/purification of proteins in Sf9 insect cells, GST pull downs, immunoprecipitation, tissue culture, transfection, luciferase assay, RT-PCR, ChIP, NMR spectroscopy/structure determination, fluorescence polarization, chromatin-based *in vitro* transcription, RNAi in *C. elegans* and fatty acid/lipid analysis by gas chromatography/mass spectrometry and thin liquid chromatography are given in Supplementary Information.

Received 7 April; accepted 1 June 2006.

Published online 21 June 2006.

1. Brown, M. S. & Goldstein, J. L. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* **89**, 331–340 (1997).
2. Eberle, D., Hegarty, B., Bossard, P., Ferre, P. & Foulfelle, F. SREBP transcription factors: master regulators of lipid homeostasis. *Biochimie* **86**, 839–848 (2004).
3. Näär, A. M. *et al.* Chromatin, TAFs, and a novel multiprotein coactivator are required for synergistic activation by Sp1 and SREBP-1a *in vitro*. *Genes Dev.* **12**, 3020–3031 (1998).
4. Näär, A. M. *et al.* Composite coactivator ARC mediates chromatin-directed transcriptional activation. *Nature* **398**, 828–832 (1999).
5. Oliner, J. D., Andresen, J. M., Hansen, S. K., Zhou, S. & Tjian, R. SREBP transcriptional activity is mediated through an interaction with the CREB-binding protein. *Genes Dev.* **10**, 2903–2911 (1996).
6. Kniazeva, M., Crawford, Q. T., Seiber, M., Wang, C. Y. & Han, M. Monomethyl branched-chain fatty acids play an essential role in *Caenorhabditis elegans* development. *PLoS Biol.* **2**, E257 (2004).
7. McKay, R. M., McKay, J. P., Avery, L. & Graff, J. M. *C. elegans*: a model for exploring the genetics of fat storage. *Dev. Cell* **4**, 131–142 (2003).
8. Jump, D. B. Dietary polyunsaturated fatty acids and regulation of gene transcription. *Curr. Opin. Lipidol.* **13**, 155–164 (2002).
9. Maxfield, F. R. & Tabas, I. Role of cholesterol and lipid organization in disease. *Nature* **438**, 612–621 (2005).
10. Moller, D. E. & Kaufman, K. D. Metabolic syndrome: a clinical and molecular perspective. *Annu. Rev. Med.* **56**, 45–62 (2005).
11. Toth, J. I., Datta, S., Athanikar, J. N., Freedman, L. P. & Osborne, T. F. Selective coactivator interactions in gene activation by SREBP-1a and -1c. *Mol. Cell. Biol.* **24**, 8288–8300 (2004).
12. Novatchkova, M. & Eisenhaber, F. Linking transcriptional mediators via the GACKIX domain super family. *Curr. Biol.* **14**, R54–R55 (2004).
13. Kato, Y., Habas, R., Katsuyama, Y., Näär, A. M. & He, X. A component of the ARC/Mediator complex required for TGF- β /Nodal signalling. *Nature* **418**, 641–646 (2002).
14. Yang, F., DeBeaumont, R., Zhou, S. & Näär, A. M. The activator-recruited cofactor/Mediator coactivator subunit ARC92 is a functionally important target of the VP16 transcriptional activator. *Proc. Natl Acad. Sci. USA* **101**, 2339–2344 (2004).
15. Chirivia, J. C. *et al.* Phosphorylated CREB binds specifically to the nuclear protein CBP. *Nature* **365**, 855–859 (1993).
16. Dai, P. *et al.* CBP as a transcriptional coactivator of c-Myb. *Genes Dev.* **10**, 528–540 (1996).
17. Radhakrishnan, I. *et al.* Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* **91**, 741–752 (1997).
18. Zor, T., De Guzman, R. N., Dyson, H. J. & Wright, P. E. Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. *J. Mol. Biol.* **337**, 521–534 (2004).
19. Parker, D. *et al.* Analysis of an activator:coactivator complex reveals an essential role for secondary structure in transcriptional activation. *Mol. Cell* **2**, 353–359 (1998).
20. Parker, D. *et al.* Role of secondary structure in discrimination between constitutive and inducible activators. *Mol. Cell. Biol.* **19**, 5601–5607 (1999).
21. Goto, N. K., Zor, T., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. *J. Biol. Chem.* **277**, 43168–43174 (2002).
22. De Guzman, R. N., Goto, N. K., Dyson, H. J. & Wright, P. E. Structural basis for cooperative transcription factor binding to the CBP coactivator. *J. Mol. Biol.* **355**, 1005–1013 (2006).
23. Bourbon, H. M. *et al.* A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II. *Mol. Cell* **14**, 553–557 (2004).
24. Ashrafi, K. *et al.* Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* **421**, 268–272 (2003).
25. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
26. Tabor, D. E., Kim, J. B., Spiegelman, B. M. & Edwards, P. A. Identification of conserved cis-elements and transcription factors required for sterol-regulated transcription of stearoyl-CoA desaturase 1 and 2. *J. Biol. Chem.* **274**, 20603–20610 (1999).
27. Watts, J. L. & Browse, J. A palmitoyl-CoA-specific delta9 fatty acid desaturase from *Caenorhabditis elegans*. *Biochem. Biophys. Res. Commun.* **272**, 263–269 (2000).
28. Watts, J. L. & Browse, J. Genetic dissection of polyunsaturated fatty acid synthesis in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **99**, 5854–5859 (2002).
29. Taubert, S., Van, M. R., Hansen, M. & Yamamoto, K. R. A Mediator subunit, MDT-15, integrates regulation of fatty acid metabolism by NHR-49-dependent and -independent pathways in *C. elegans*. *Genes Dev.* **20**, 1137–1149 (2006).
30. Yuan, C. X., Ito, M., Fondell, J. D., Fu, Z. Y. & Roeder, R. G. The TRAP220 component of a thyroid hormone receptor-associated protein (TRAP) coactivator complex interacts directly with nuclear receptors in a ligand-dependent fashion. *Proc. Natl Acad. Sci. USA* **95**, 7939–7944 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We were unable to cite many original papers owing to space constraints. We acknowledge G. Gill, N. Dyson and B. Spiegelman for comments on the manuscript. We thank B. Lüscher, R. Sordella and M. Classon for reagents, N. J. Moerke for his assistance with the fluorescence polarization experiments, and M. Dedmon for providing purified ARC105. A.M.N. is a Dammern Scholar of the Damon Runyon Cancer Research Foundation. This work was supported by the NIH, the Damon Runyon Cancer Research Foundation, and the Milton Foundation of Harvard University.

Author Information Atomic coordinates of the amino-terminal region of ARC105 containing the KIX domain has been deposited in the Protein Data Bank with the accession number 2GUT. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.M.N. (naar@helix.mgh.harvard.edu).

Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice

Kenong Xu¹, Xia Xu¹, Takeshi Fukao², Patrick Canlas¹, Reyce Maghirang-Rodriguez³, Sigrid Heuer³, Abdelbagi M. Ismail³, Julia Bailey-Serres², Pamela C. Ronald¹ & David J. Mackill³

Most *Oryza sativa* cultivars die within a week of complete submergence—a major constraint to rice production in south and southeast Asia that causes annual losses of over US\$1 billion and affects disproportionately the poorest farmers in the world^{1,2}. A few cultivars, such as the *O. sativa* ssp. *indica* cultivar FR13A, are highly tolerant and survive up to two weeks of complete submergence owing to a major quantitative trait locus designated *Submergence 1* (*Sub1*) near the centromere of chromosome 9 (refs 3–6). Here we describe the identification of a cluster of three genes at the *Sub1* locus, encoding putative ethylene response factors. Two of these genes, *Sub1B* and *Sub1C*, are invariably present in the *Sub1* region of all rice accessions analysed. In contrast, the presence of *Sub1A* is variable. A survey identified two alleles within those *indica* varieties that possess this gene: a tolerance-specific allele named *Sub1A-1* and an intolerance-specific allele named *Sub1A-2*. Overexpression of *Sub1A-1* in a submergence-intolerant *O. sativa* ssp. *japonica* conferred enhanced tolerance to the plants, downregulation of *Sub1C* and upregulation of *Alcohol dehydrogenase 1* (*Adh1*), indicating that *Sub1A-1* is a primary determinant of submergence tolerance. The FR13A *Sub1* locus was introgressed into a widely grown Asian rice cultivar using marker-assisted selection. The new variety maintains the high yield and other agronomic properties of the recurrent parent and is tolerant to submergence. Cultivation of this variety is expected to provide protection against damaging floods and increase crop security for farmers.

Submergence of plants inhibits aerobic respiration and photosynthesis, and stimulates a variety of responses that can enhance survival, such as a switch from aerobic to anaerobic respiration⁷. In contrast to deep-water rice cultivars that avoid submergence stress by growing above the water surface and thereby restoring gas exchange⁸, submergence-tolerant rice can survive 10–14 days of complete submergence and renew growth when the water subsides⁹—although the duration of survival is also influenced by environmental factors such as water turbidity, temperature and light levels¹⁰. The *Sub1* locus was mapped to an interval of 0.06 centimorgans on chromosome 9 using a mapping population (DX202) of 4,022 plants developed from the hybridization of a tolerant *indica* derivative of the FR13A cultivar (IR40931-26) and the intolerant *japonica* cultivar M-202 (Supplementary Fig. 1a, b; Supplementary Tables 1–3). Physical coverage of this region was obtained with five overlapping bacterial artificial chromosome (BAC) clones derived from submergence-intolerant *indica* rice varieties and a nearly complete contig of 13 binary clones from IR40931-26 (Supplementary Fig. 1b). The *Sub1* region, bordered by the markers CR25K and SSR1A, physically spans over 182 kilobases (kb). This interval encodes three genes containing ethylene-response-factor (ERF) domains and designated *Sub1A*, *Sub1B* and *Sub1C*, ten non-ERF genes including four transcribed and six

hypothetical protein-coding genes, and >50% retrotransposon-related sequences (Fig. 1a; Supplementary Fig. 1; Supplementary Table 4). The corresponding region of the *japonica* genome represented by the sequenced variety Nipponbare spans 142 kb and is considerably rearranged. Notably, *Sub1A* is absent from the Nipponbare genome¹¹. Recombination was suppressed in this region in the mapping population, as revealed by the 10.7-fold higher-than-average recombination ratio (3,030 kb cM⁻¹ in the *Sub1* region versus 282 kb cM⁻¹ for the entire genome)^{5,12}. This could reflect the proximity of the *Sub1* locus to the centromere and/or the presence of genomic rearrangements that have altered continuity in this region in the two rice subspecies¹³.

Plant proteins that contain ERF domains are known regulators of abiotic and biotic stress responses^{14,15}. The accumulation of *Sub1A* and *Sub1C* messenger RNAs was strongly but transiently promoted by submergence and further reduced on de-submergence in seedling leaves of tolerant FR13A (Fig. 1b). *Sub1C* mRNA induction was earlier and more pronounced in intolerant Nipponbare compared with FR13A (Fig. 1b), suggesting that the rapid induction of *Sub1A* limits expression of *Sub1C*. *Adh1* gene transcript levels were more strongly induced in the tolerant line, indicating that *Sub1A* may

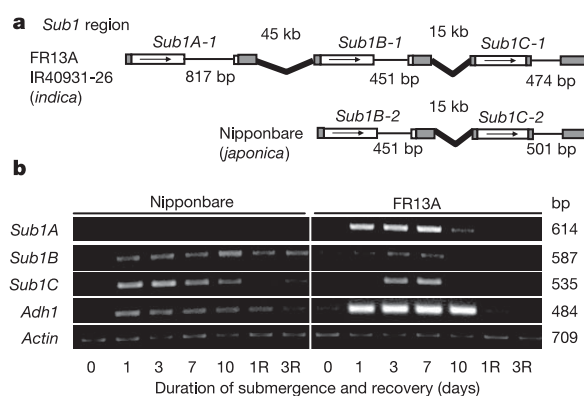


Figure 1 | *Sub1* region gene composition and submergence-induced mRNA accumulation in rice. **a**, ERF gene structure and organization in tolerant *indica* (IR40931-26) and intolerant *japonica* (Nipponbare). Arrows: direction of transcription; shaded boxes: untranslated regions; open boxes: coding sequence; thin lines: introns; thick lines: intergenic regions. Gene structure was determined by comparison of genomic and cDNA sequences of *Sub1B* (AK106057, AK068688) and *Sub1C* (AK060090, AK072749). **b**, Semi-quantitative RT-PCR assessment of gene transcript levels in shoot tissue from tolerant (FR13A) and intolerant (Nipponbare) genotypes following 1–10-d submergence and a subsequent 1–3-d recovery (1R and 3R, respectively).

¹Department of Plant Pathology, University of California, Davis, California 95616, USA. ²Center for Plant Cell Biology, Department of Botany and Plant Sciences, University of California, Riverside, California 92521-0124, USA. ³International Rice Research Institute, DAPO Box 7777, Metro Manila, The Philippines.

positively regulate certain acclimation responses (Fig. 1b). In contrast, *Sub1B* transcripts increased only slightly during submergence (Fig. 1b; Supplementary Fig. 2a). The ten non-ERF genes in the *indica Sub1* region showed no evidence of expression in seedling leaves before or during submergence in IR40931-26 or the intolerant variety M-202 (data not shown). These results indicated that the three ERF-domain-containing genes, particularly *Sub1A*, were strong candidates for the genetic determinant controlling submergence tolerance.

The three SUB1 proteins each possess a single copy of the 56-amino-acid DNA-binding domain characteristic of the ERF subfamily of the plant *Apetala2*-like transcription factors (Supplementary Fig. 3). SUB1B and SUB1C have the two signature amino acids—alanine and aspartic acid—at positions 13 and 18, respectively, of the ERF domain that are characteristic of the B2 subgroup of ERF proteins¹⁶. In contrast, SUB1A has a serine at position 13, as found in two members of the B1b subgroup¹⁵, but falls in the same clade as the other two SUB1 proteins¹⁴. Within the ERF domain, SUB1A shares 87.7% and 77.2% sequence identity with SUB1B and SUB1C, respectively (Supplementary Fig. 3). A full-length *Sub1A* transcript of 1312 nucleotides was obtained from mRNA of submerged IR40931-26 plants by reverse transcription–polymerase chain reaction (RT–PCR). The transcript consists of a 5' untranslated region (UTR) of 149 nt, an open reading frame (ORF) of 846 nucleotides encoding a deduced protein of 281 amino acids, and a 3' UTR of 317 nucleotides (Fig. 1a; Supplementary Fig. 4). The 5' end of the *Sub1A* complementary DNA includes 30 nucleotides of ORF not predicted by gene-prediction algorithms. The *Sub1A* gene has one intron of 817 base pairs (bp) and two exons, the second exon comprising only 11 bp including the stop codon. A similar genomic structure was found for *Sub1B* (second exon 15 bp), whereas *Sub1C* has only one intron, located in the 3' UTR (Fig. 1a).

A survey of *Sub1* locus haplotypes in 17 *indica* and four *japonica* varieties identified two *Sub1A*, nine *Sub1B* and seven *Sub1C* alleles on the basis of variation in amino-acid sequence (Table 1; Supplementary Figs 3, 5 and 6). The *Sub1A-1* and *Sub1C-1* alleles are limited to all six submergence-tolerant accessions, three of which were independently isolated submergence-tolerant varieties, including FR13A. There was no *Sub1B* allele identified as being specific to submergence tolerance. Variations in putative mitogen-activated protein kinase (MAPK) sites distinguish the tolerant and intolerant alleles of *Sub1A* and *Sub1C*. In the tolerant *Sub1A-1* allele a single nucleotide polymorphism at position 556 is responsible for a Pro 186 (intolerant) to Ser 186 (tolerant) substitution in a MAPK site (PXS/TP, where “X” is any amino acid¹⁷). Conversely, the *Sub1C-1* allele of tolerant lines lacks a MAPK phosphorylation site (underlined) present in the alleles of the intolerant accessions (tolerant: SPPP₁₇₅PEQPAAPV; intolerant: SL/PPT₁₇₅PPPP/E(P)_{0–3}; (where P can be 0 or 3 in intolerant rice varieties) Table 1; Supplementary Figs 3 and 6).

These potential phosphorylation sites are located in variable regions immediately carboxy-terminal to the ERF domain and may be of significance as phosphorylation can modulate DNA binding by ERF proteins^{14,18}.

The allelic survey further revealed that *Sub1A* is absent from five out of seventeen *indica* varieties and all four *japonica* varieties analysed, including the fully sequenced genome of Nipponbare¹¹ (Table 1). Assay of gene expression in a selection of *indica* varieties with *Sub1A* revealed that submergence tolerance is correlated with possession of the strongly submergence-induced *Sub1A-1* and intolerance is associated with the poorly submergence-induced *Sub1A-2* or complete absence of this gene (Fig. 1b; Supplementary Fig. 2a). Nucleotide polymorphisms in the *Sub1A* alleles, including the 5'-flanking regions (Supplementary Fig. 4), could be responsible for their differential expression. In contrast, submergence tolerance is correlated with limited induction of *Sub1C* under submergence, whereas intolerance is associated with high levels of *Sub1C* mRNA (Fig. 1b; Supplementary Fig. 2a). Together, these data suggested that stable transformation of *japonica* rice with an ectopically expressed *Sub1A-1* would downregulate *Sub1C* and confer submergence tolerance.

To test this hypothesis, we transformed the intolerant *japonica* variety Liaogeng with a *Sub1A-1* full-length cDNA under the control of the maize *Ubiquitin1* promoter^{19,20} (*Ubi:Sub1A*). A screen of seedlings after 11 days of submergence identified four T₁ families, derived from independent T₀ *Ubi:Sub1A*⁺ lines, with submergence-tolerant transgenic individuals, and progeny from two families were examined in detail. T₁ families 1 and 3 showed a correlation between high expression of the *Sub1A-1* transgene and submergence tolerance (Fig. 2c; Supplementary Fig. 2b). As observed in the FR13A descendant IR40931-26, tolerant *Sub1A-1*⁺ plants showed a significant impairment of shoot elongation under submergence compared with the intolerant parent Liaogeng and non-transgenic siblings (Fig. 2a, b; Supplementary Fig. 2c). *Ubi:Sub1A-1* conferred a pleiotropic phenotype including reduced plant height under normal and submergence conditions (Fig. 2a, b) and enhanced expression of *Adh1* under normal growth conditions (Fig. 2c). Under submergence, the transgenic progeny showed reduced *Sub1C* (Fig. 2d) and enhanced *Adh1* (Supplementary Fig. 2d) mRNA accumulation concomitant with increased survival, as characteristic of tolerant *indica* varieties (Fig. 1b; Supplementary Fig. 2a). Although ectopic expression of *Ubi:Sub1A-1* reduced plant height, submergence tolerance was independent of plant height at the time of inundation (Fig. 2a; Supplementary Fig. 2c). It is known that entrapment of ethylene during submergence leads to decreased abscisic acid levels and increased gibberellin sensitivity, and has ramifications on cellular metabolism as well as stem and leaf elongation^{10,21,22}. *Sub1* genotype controls ethylene and gibberellin mediated changes in gene expression including regulation of genes that control carbohydrate consumption and

Table 1 | Haplotypes of the *Sub1* locus based on alleles of the ERF-like genes in rice varieties

Line or cultivar	Submergence phenotype	Subspecies	<i>Sub1A</i> allele	<i>Sub1B</i> allele	<i>Sub1C</i> allele
FR13A, IR40931-26, DX18-121, IR48930	Tolerant	<i>indica</i>	A-1	B-1	C-1
Goda Heenati	Tolerant	<i>indica</i>	A-1	B-6	C-1
Kurkaruppan	Tolerant	<i>indica</i>	A-1	B-3	C-1
LMNIII	ND	<i>indica</i>	A-2	B-1	C-4
Teqing, CO39, IR64, IR64-M6D6-933-1-2, 93-11	Intolerant	<i>indica</i>	A-2	B-1, B-7	C-3, C-5
IR24, IRBB21, Swarna*	Intolerant	<i>indica</i>	Absent	B-8, B-5	C-6
IR50	Intolerant	<i>indica</i>	Absent	B-9	C-7
Habiganj aman	Intolerant	<i>indica</i>	Absent	B-4	C-6
Nipponbare, Liaogeng, M-202, Taipei309	Intolerant	<i>japonica</i>	Absent	B-2	C-2

Allele designations were based on the amino-acid sequence of the putative proteins (Supplementary Figs 3, 5 and 6). The submergence-tolerant *indica*-like variety FR13A is from Orissa, in eastern India. DX18-121 is an *indica/japonica* hybrid derivative. The submergence-tolerant varieties Kurkaruppan and Goda Heenati are from Sri Lanka. IR48930, IR40931-26 and DX18-121 are derivatives of FR13A. The primary locus conferring tolerance in FR13A and Kurkaruppan was reported to be similar but different from Goda Heenati²⁰. However, submergence tolerance in Goda Heenati is also largely controlled by the *Sub1* locus (K.X. and D.J.M. unpublished data). Molecular marker studies indicate considerable divergence between Goda Heenati and FR13A (D.J.M. unpublished data). GenBank accessions of 93-11 containing *Sub1A*, *Sub1B* and *Sub1C* are AAAA01009971, AAAA01020021 and AAAA01005744, respectively. ND, not determined. The varieties are grouped based primarily on common alleles of *Sub1A* and *Sub1C*.

* Swarna lacks *Sub1A* and its alleles of *Sub1B* and *Sub1C* were not determined.

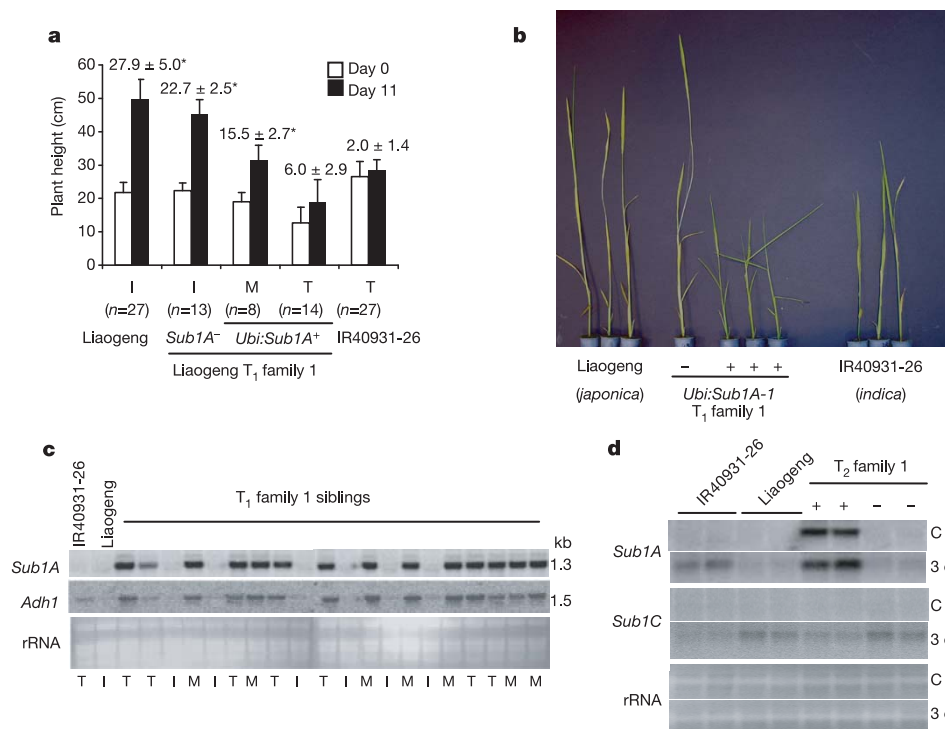


Figure 2 | Characterization of submergence response in transgenic rice ectopically expressing *Sub1A-1*. **a**, Comparison of plant height before submergence and after 11 days of submergence of the controls (non-transformed Liaogeng and IR40931-26) and siblings from T₁ family 1, including non-transgenic (*Sub1A*⁻) intolerant (I) plants, and transgenic (*Sub1A*⁺) intermediate tolerant (M) and tolerant (T) plants. At day 0 *Sub1A*⁺ plants were significantly shorter than *Sub1A*⁻ plants ($P < 0.0001$, unpaired *t*-test). Values above the bars are the mean and s.d. of the change in plant height following submergence. Error bars are also s.d. An asterisk indicates a significant increase in plant height in response to submergence at day 11 ($P < 0.0001$, unpaired *t*-test). **b**, Comparison of plant height of T₁

family 1 siblings and control plants following 11 days of submergence. **c**, Northern blot analysis of *Sub1A-1* and *Adh1* mRNA levels in shoot tissue harvested 10 days following de-submergence in T₁ family 1 siblings and control plants. Hybridization was performed with rice *Sub1A-1* and *Adh1* cDNA probes. **d**, Northern blot analysis of *Sub1A* and *Sub1C* mRNA levels in shoot tissue harvested immediately following submergence for 3 days or in control non-submerged tissue (C) from four T₂ siblings generated by self-pollination of a T₁ family 1 hemizygote. Siblings with the *Ubi:Sub1A* transgene (+); siblings lacking the transgene (-). Hybridizations were to *Sub1A-1* cDNA and *Sub1C-1* 3'-UTR probes, respectively.

cell elongation²³. A genetic interaction between *Sub1A* and *Sub1C* may be of relevance because antagonistic relationships between ERFs have been recognized^{15,24,25}. Detailed analyses of the function of *Sub1A* and *Sub1C* in the submergence response by targeted RNA interference (RNAi) and overexpression constructs are ongoing.

We used polymorphic molecular markers for *Sub1* with markers that flanked the locus to introgress the *Sub1* genes into the widely grown Indian variety Swarna, which lacks *Sub1A*. *Sub1* markers were used for selection of tolerant progenies, in combination with 5–12 background markers for each of the 12 rice chromosomes^{26–28}. Marker assisted selection (MAS) application in the first backcross generation (BC₁F₁) was used to identify individual plants with the fewest IR49830-7-1-2-3 (FR13A descendant) chromosomal segments. Selected BC₁F₁ plants were used to generate BC₂F₃ and BC₃F₂ Swarna-*Sub1* lines that were genotypically identical to Swarna, except for the *Sub1* haplotype and adjacent markers. Both Swarna-*Sub1* lines showed strong submergence tolerance (Fig. 3). Field trials with the BC₂F₃ plants grown under control (non-submerged) conditions in the Philippines indicated no differences between the two varieties in terms of yield (Swarna: 6.3 ± 0.1 t ha⁻¹; Swarna-*Sub1*: 6.4 ± 0.1 t ha⁻¹), plant height (Swarna: 105 ± 1.4 cm; Swarna-*Sub1*: 106 ± 1.2 cm), harvest index (both 0.35) and grain quality as indicated by amylose content (Swarna: 26.4%; Swarna-*Sub1*: 25.9%). Development of submergence-tolerant varieties using these procedures is at an advanced stage for Laos, Bangladesh and India, and has already been reported in Thailand²⁹.

The results presented here confirm that submergence tolerance is conferred by a haplotype of the complex *Sub1* locus, with ectopic



Figure 3 | Introgression of the FR13A *Sub1* haplotype into an intolerant variety by MAS confers submergence tolerance. The *Sub1* region donor line IR49830 (an FR13A derivative) was introduced into the submergence-intolerant *indica* variety Swarna by backcrossing (BC) with MAS using markers for the *Sub1* region (SSR1, RM316, RM464, RM464A, RM219 and RM524) and the 12 chromosomes^{25–27}. Individual F₁ plants were selected from BC₁, BC₂ and BC₃ that carried the FR13A *Sub1* haplotype with the least IR49830 background. Fourteen-day-old seedlings were submerged for 14 days and photographed 14 d after de-submergence.

Sub1A-1 expression being sufficient to enhance tolerance. The finding of the identical *Sub1* haplotype in accessions from submergence-prone areas in Sri Lanka and eastern India (Table 1) suggests that rice grains from submergence-tolerant plants may have been transported over 1,000 km and subsequently introgressed into local varieties, further indicating the agronomic importance of the *Sub1* locus.

METHODS

Sub1 characterization. Details of rice genotypes, growth and treatment conditions, and mapping are provided in Supplementary Methods. An F₂ mapping population of 4,022 plants expanded from DX202 (ref. 4), derived from a cross between DX18-121 (a tolerant line derived from *indica* IR40931-26, a descendant of FR13A) and the intolerant *japonica* cultivar M-202, was used. Submergence treatment was conducted as described previously⁴. The fine-mapping of the *Sub1* locus was accomplished with 24 markers specific to the *Sub1* region. For gene expression analyses, total RNA was isolated from seedling leaves and analysed as detailed in Supplementary Methods.

Generation of submergence-tolerant rice. To overexpress *Sub1A-1*, a binary construct *Ubi:Sub1A-1-C1300* carrying the full-length *Sub1A-1* cDNA was transformed into an intolerant *japonica* cultivar, Liaogeng, using *Agrobacterium tumefaciens* (EHA105) as described in Chern *et al.*²⁰. Integration of the *Ubi:Sub1A-1* transgene was verified by PCR using a maize *Ubiquitin1* promoter-specific primer and a *Sub1A* specific primer. Submergence tolerance and gene expression in the transgenic rice was evaluated as described in Supplementary Methods. Swarna-*Sub1* was produced by crossing the Indian variety Swarna to the FR13A descendant IR49830-7-1-2-3, followed by subsequent backcrossing to Swarna. MAS of progeny with polymorphic markers was performed by PCR analysis of genomic DNA.

Received 23 March; accepted 15 May 2006.

- Dey, M. & Upadhyaya, H. in *Rice Research in Asia: Progress and Priorities* (eds Evenson, R., Herdt, R. & Hossain, M.) 291–303 (CAB International, Wallingford, 1996).
- Herdt, R. W. in *Rice Biotechnology* (eds Khush, G. S. & Toenniessen, G. H.) 19–54 (CAB International, Wallingford, 1991).
- Xu, K. & Mackill, D. J. A major locus for submergence tolerance mapped on rice chromosome 9. *Mol. Breed.* **2**, 219–224 (1996).
- Xu, K., Xu, X., Ronald, P. C. & Mackill, D. J. A high-resolution linkage map in the vicinity of the rice submergence tolerance locus *Sub1*. *Mol. Gen. Genet.* **263**, 681–689 (2000).
- Harushima, Y. *et al.* A high-density rice genetic linkage map with 2275 markers using a single F₂ population. *Genetics* **148**, 479–494 (1998).
- Chen, M. *et al.* An integrated physical and genetic map of the rice genome. *Plant Cell* **14**, 537–545 (2002).
- Fukao, T. & Bailey-Serres, J. Plant responses to hypoxia—is survival a balancing act? *Trends Plant Sci.* **9**, 449–456 (2004).
- Sauter, M. Rice in deep water: “How to take heed against a sea of troubles”. *Naturwissenschaften* **87**, 289–303 (2000).
- Catling, D. *Rice in Deep Water* 5 (Macmillan, London, 1992).
- Ram, P. C. *et al.* Submergence tolerance in rainfed lowland rice: physiological basis and prospects for cultivar improvement through marker-aided breeding. *Field Crops Res.* **76**, 131–152 (2002).
- International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* **436**, 793–800 (2005).
- Arumanagathan, K. & Earle, E. D. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9**, 229–241 (1991).
- Wu, J. *et al.* A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell* **14**, 525–535 (2002).
- Gutterson, N. & Reuber, T. L. Regulation of disease resistance pathways by AP2/ERF transcription factors. *Curr. Opin. Plant Biol.* **7**, 465–471 (2004).
- McGrath, K. C. *et al.* Repressor- and activator-type ethylene response factors functioning in jasmonate signaling and disease resistance identified via a genome-wide screen of *Arabidopsis* transcription factor gene expression. *Plant Physiol.* **139**, 949–959 (2005).
- Magnani, E., Sjolander, K. & Hake, S. From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. *Plant Cell* **16**, 2265–2277 (2004).
- Parnell, S. C. *et al.* Phosphorylation of the RGS protein Sst2 by the MAP kinase Fus3 and use of Sst2 as a model to analyse determinants of substrate sequence specificity. *Biochemistry* **44**, 8159–8166 (2005).
- Cheong, Y. H. *et al.* BWMK1, a rice mitogen-activated protein kinase, locates in the nucleus and mediates pathogenesis-related gene expression by activation of a transcription factor. *Plant Physiol.* **132**, 1961–1972 (2003).
- Christensen, A. H. & Quail, P. H. Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Res.* **5**, 213–218 (1996).
- Chern, M.-S., Fitzgerald, H. A., Canlas, P. E., Navarre, D. A. & Ronald, P. C. Overexpression of a rice NPR1 homolog leads to constitutive activation of defense response and hypersensitivity to light. *Mol. Plant Microbe Interact.* **18**, 511–520 (2005).
- Kende, H., van der Knaap, E. & Cho, H.-T. Deepwater rice: a model plant to study stem elongation. *Plant Physiol.* **118**, 1105–1110 (1998).
- Jackson, M. B. & Ram, P. C. Physiological and molecular basis of susceptibility and tolerance of rice plants to complete submergence. *Ann. Bot.* **91**, 227–241 (2003).
- Fukao, T., Xu, K., Ronald, P. & Bailey-Serres, J. A variable cluster of ethylene responsive-like factors regulates metabolic and developmental acclimation responses to submergence in rice. *Plant Cell* (in the press).
- Fujimoto, S. Y., Ohta, M., Usui, A., Shinshi, H. & Ohme-Takagi, M. *Arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell* **12**, 393–404 (2000).
- Song, C.-P. *et al.* Role of an *Arabidopsis* AP2/EREBP-type transcriptional repressor in abscisic acid and drought stress responses. *Plant Cell* **17**, 2384–2396 (2005).
- Chen, X., Temnykh, S., Xu, Y., Cho, Y. G. & McCouch, S. R. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **95**, 553–567 (1997).
- Temnykh, S. *et al.* Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **100**, 697–712 (2000).
- Xu, K., Deb, R. & Mackill, D. J. A microsatellite marker and a codominant PCR-based marker for marker-assisted selection of submergence tolerance in rice. *Crop Sci.* **44**, 248–253 (2004).
- Siangliw, M., Toojinda, T., Tragoonrun, S. & Vanavichit, A. Thai jasmine rice carrying QTLch9 (*Sub1*) is submergence tolerant. *Ann. Bot.* **91**, 255–261 (2003).
- Setter, T. L. *et al.* Physiology and genetics of submergence tolerance in rice. *Ann. Bot.* **79**, 67–77 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Dvorak for discussions of experimental design, H. Zhang for providing BAC clones, D. Chen, J. S. Jeon, J. Ni, P. Colowit, D. Ruan, R. Bruskewich, G. Vergara, M. Nas, A. Pamplona, C. N. Neeraja, and S. Singh for their technical assistance and suggestions, support and discussions, and the IRRI Grain Quality, Nutrition, and Postharvest Center for amylose data. This work was supported by USDA-NRIGP grants to D.J.M. and P.C.R., and to J.B.-S. and P.C.R., and USAID Linkage Project funds to J.B.-S. and P.C.R. The work on transferring *Sub1* to Swarna was supported by a grant from the German Federal Ministry for Economic Cooperation and Development (BMZ) to D.J.M. and A.M.I.

Author Information Sequences were submitted to GenBank/EMBL/DDJB under accession numbers DQ011597–DQ011607 and DQ453964–DQ453966. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.C.R. (pcronald@ucdavis.edu) or D.J.M. (d.mackill@cgiar.org).

Assembly dynamics of microtubules at molecular resolution

Jacob W. J. Kerssemakers^{1,2}, E. Laura Munteanu¹, Liedewij Laan¹, Tim L. Noetzel², Marcel E. Janson^{1,3} & Marileen Dogterom¹

Microtubules are highly dynamic protein polymers¹ that form a crucial part of the cytoskeleton in all eukaryotic cells. Although microtubules are known to self-assemble from tubulin dimers, information on the assembly dynamics of microtubules has been limited, both *in vitro*^{2,3} and *in vivo*^{4,5}, to measurements of average growth and shrinkage rates over several thousands of tubulin subunits. As a result there is a lack of information on the sequence of molecular events that leads to the growth and shrinkage of microtubule ends. Here we use optical tweezers to observe the assembly dynamics of individual microtubules at molecular resolution. We find that microtubules can increase their overall length almost instantaneously by amounts exceeding the size of individual dimers (8 nm). When the microtubule-associated protein XMAP215 (ref. 6) is added, this effect is markedly enhanced and fast increases in length of about 40–60 nm are observed. These observations suggest that small tubulin oligomers are able to add directly to growing microtubules and that XMAP215 speeds up microtubule growth by facilitating the addition of long oligomers. The achievement of molecular resolution on the microtubule assembly process opens the way to direct studies of the molecular mechanism by which the many recently discovered microtubule end-binding proteins regulate microtubule dynamics in living cells^{7–9}.

The core structure of a microtubule (MT) consists of typically 13 or 14 protofilaments forming a hollow tube (Fig. 1a). MT assembly is accompanied by the hydrolysis of tubulin-bound GTP, which occasionally triggers the MT to undergo a ‘catastrophe’ and switch to a state of rapid disassembly¹. Information about the structure of growing and shrinking MT ends as well as the (preferred) conformational state of GTP-bound and GDP-bound tubulin in these situations comes from static electron microscopy studies^{10–12}, which indicate that growing MT ends consist of sheets of slightly outward-curved protofilaments and that shrinking MT ends consist of individual protofilaments that curve outwards more strongly.

To obtain dynamic information on the growth and shrinkage of MTs at the resolution of single tubulin dimers (8 nm) we use a technique based on optical tweezers in which we allow dynamic MT plus ends to grow and shrink against a microfabricated barrier¹³. We show results for MTs assembling from pure tubulin, and then show how, on a molecular scale, the growth process is altered by the presence of XMAP215, an evolutionarily conserved protein¹⁴ that enhances the growth rate of MTs^{15–17}. Our set-up is shown schematically in Fig. 1b (see also Supplementary Methods). MTs are nucleated by an axoneme, a naturally occurring rigid bundle of multiple stabilized MTs, to which a bead is attached near one end. The bead–axoneme construct is suspended in a ‘keyhole’ optical trap near a rigid, microfabricated barrier that was built into a flow cell.

The keyhole trap is used to control both the position of the bead and the direction of the axoneme. MT growth is initiated by flowing in a mixture of tubulin and GTP, with or without the addition of XMAP215. The conditions are chosen such that most of the time only one or two MTs are nucleated by the plus end of the axoneme (see Supplementary Information). When a MT reaches the barrier, further length increases lead to displacement of the bead in the trap. The relation between the increase in MT length and the displacement of the bead depends on the stiffness of the bead–axoneme construct, which we measure independently by pushing the barrier against the construct before growth is initiated. Displacement of the bead in the optical trap leads to an increasing restoring force on the bead (given by the trap stiffness k_{trap} multiplied by the bead displacement) that pushes the growing MT tip against the barrier. In all of the analysis below, we keep the growing MT short (less than 1 μm) to prevent the MT from buckling under this compressive load¹⁸.

Figure 1c shows the restoring force exerted by the trap during a

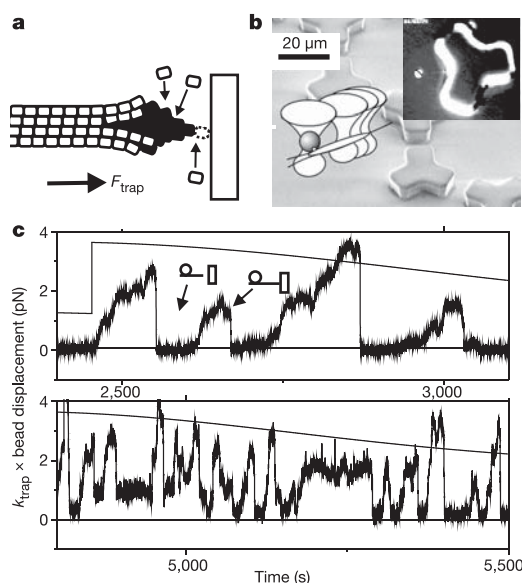


Figure 1 | Measuring growth dynamics of MTs with optical tweezers. **a**, Schematic view of a growing MT. **b**, Schematic and DIC image of a ‘keyhole’ optical trap holding a bead–axoneme construct in front of a microfabricated barrier. **c**, Growth and shrinkage events of individual MTs in the absence (upper panel) and presence (lower panel) of XMAP215. The smooth curves give estimates of the gradually decreasing protein concentration (maximum tubulin concentration 20 μM).

¹Foundation for Fundamental Research on Matter (FOM) Institute for Atomic and Molecular Physics (AMOLF), Kruislaan 407, 1098 SJ Amsterdam, The Netherlands.

²Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG) Dresden, Pfotenhauerstrasse 108, 01307 Dresden, Germany. ³University of Pennsylvania, Department of Cell and Developmental Biology, 421 Curie Boulevard, Philadelphia, Pennsylvania 19104-6058, USA.

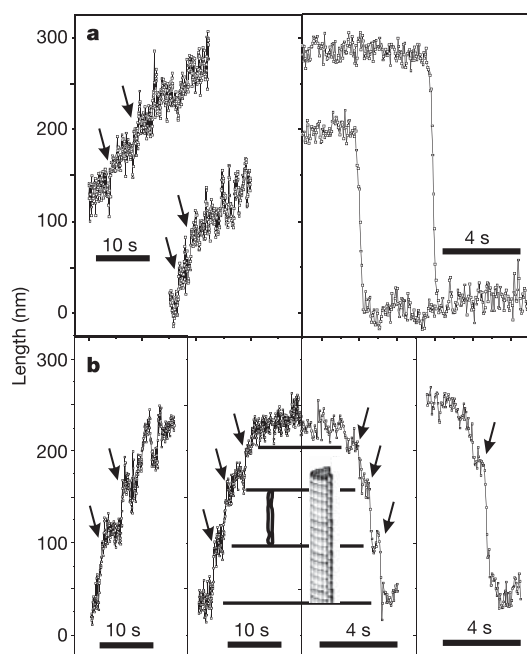


Figure 2 | High-resolution details of growth and shrinkage events. Length traces are shown for events without (a) and with (b) XMAP215. Arrows indicate fast length changes (steps) that are distinguishable from the noise by eye. During subsequent growth and shrinkage events in the presence of XMAP215, steps sometimes appear in register with each other (b, middle, lines). The insets show schematic pictures of an XMAP215 molecule (left) and a MT (right) at the same vertical scale as the length data.

sequence of MT growth and shrinkage events. The upper panel shows regular MT growth, with an initial tubulin concentration of 20 μM . The protein concentration slowly decreases as indicated by the smooth curves (as the result of a slow buffer flow). Growth often

comes to an apparent halt at a few piconewtons of force before catastrophes occur, which is consistent with what is known for the effect of force on the assembly dynamics of MT plus ends^{18–20}. The lower panel shows growth from the same construct after replacement of the surrounding solution with 20 μM tubulin and 150 nM XMAP215. With XMAP215 present, growth is faster and catastrophes occur more frequently, in agreement with previous reports^{15–17,21} and our own observations in the absence of force (Supplementary Information). The frequent occurrence of catastrophes confirms that the observed growth is that of a plus end of a MT.

Figure 2 shows part of the same data at higher resolution, this time with the MT length on the y axis. These data gives us unprecedented resolution on the growth dynamics of individual MTs, limited only by the thermal noise on the bead in the trap and the stiffness of the bead–axoneme construct (about 5–10 nm root-mean-square in practice). In these data we occasionally observe step-like length increases that are clearly distinguishable from the experimental noise, even by eye (Fig. 2a, arrows). When we add XMAP215 (Fig. 2b), fast increases occur more frequently and are also larger, on the order of several tens of nanometres, occurring on a timescale of about 100 ms or less (the average MT growth rate under these conditions corresponds to at most 1 or 2 nm per 100 ms). In the presence of XMAP215, steps are also clearly observed during shrinkage events (Fig. 2b, right). These steps are of similar sizes to those observed during growth, sometimes even in marked registry with previous growth steps (Fig. 2b, middle). The sizes of the large steps often seem close to the known length of the XMAP215 protein itself²², as indicated in Fig. 2b. In addition to steps we observe periods with more gradual increases in length, in which steps are not clearly distinguishable. Given our experimental resolution this is to be expected: MT ends are likely to have multiple binding sites for tubulin dimers at unequal positions along the MT axis (see Fig. 1a), and growth due to single dimer additions, for example, should lead to mostly small steps with a maximum of 8 nm.

To be able to analyse in an unbiased way the steps observed by eye,

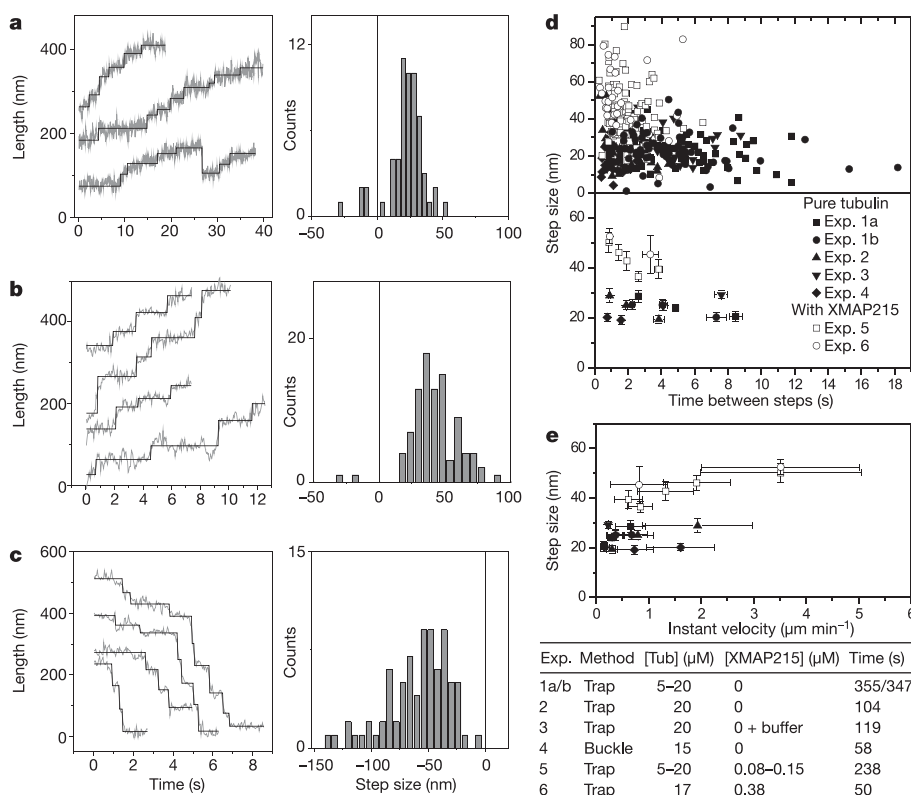


Figure 3 | Quantifying the sizes of the large steps with our step-fitting algorithm. Step fits and associated step histograms are shown for regular MT growth (a, experiment 1a), growth with XMAP215 present (b, experiment 5) and shrinkage with XMAP215 present (c, experiment 5). d, Step sizes for six different growth experiments with conditions and analysed growth times as listed in the table (experiments 1a, 1b and 5 were performed with the same bead–axoneme construct; shrinkage was analysed only for experiment 5 (63 s); experiment 4 involves data obtained previously with a different ‘buckling’ method²⁰). Top: individual step sizes plotted against the time between steps. Bottom: the same data averaged over 20 steps (the last data point may contain fewer steps). e, Averaged step sizes as in d but plotted against ‘instant velocity’ (step size divided by time between steps). Error bars in d and e represent s.e.m.

we developed a step-fitting algorithm as outlined in the Supplementary Methods. Because in our system many of the steps were smaller than the noise, we could use the algorithm only to find and quantify the largest steps (Fig. 3). In the histograms presented, any steps (or lack thereof) close to or below the noise level are therefore insignificant. For pure tubulin growth, the step fit result and the associated histogram (Fig. 3a) confirm the presence of steps up to 20–30 nm in size. This is significantly larger than the tubulin dimer size of 8 nm. In the presence of XMAP215, the histograms for growth (Fig. 3b) and shrinkage (Fig. 3c) show larger step-like changes, around 40–60 nm in size. The observed step sizes are independent of how quickly steps follow each other (that is, the rate at which MTs grow). Our experiments were performed at various tubulin concentrations and forces (both of which affect the average MT growth velocity²⁰), as well as for two different XMAP215/tubulin ratios. When we plot individual (Fig. 3d, top) and averaged (Fig. 3d, bottom) step sizes as a function of the time between steps for six different experiments (see the table in Fig. 3e), we find that the step sizes are always larger in the presence of XMAP215, even when the resulting (instant) growth velocity is the same (plotted in Fig. 3e).

The observation of fast increases in length larger than the dimer size might have two interpretations: either we were observing extended closure events of outward-curved sheets existing at the ends of growing MTs, or we were observing the addition of tubulin oligomers larger than individual tubulin dimers. To estimate the expected length increase in these two cases, we must consider the geometry and rigidity of tubulin oligomers and MT sheets as well as

the distortions that one expects in the presence of a few piconewtons of force. In our experiments MTs grew relatively slowly and never very long (typically a few hundred nanometres). In Fig. 4a we show schematically a reasonable estimate, based on electron microscopy studies^{10,23}, of a typical MT sheet under these conditions, drawn to scale. The estimates shown in Fig. 4 (see also Supplementary Methods) indicate that length increases due to oligomer addition, as opposed to sheet closure events, should be readily detectable.

Our observations therefore indicate that MT assembly might not always occur simply by the addition of individual tubulin dimers. Rather, small oligomers of up to three tubulin dimers seem to be able to attach to growing MTs as well. This is consistent with observations that MT assembly is reduced when tubulin oligomers are centrifuged away from MT polymerization solutions¹². The step sizes of 40–60 nm in the presence of XMAP215 indicate that XMAP215 might facilitate the addition of even longer oligomers. Again, there are two ways in which this could be accomplished^{24,25}. First, XMAP215 could template the assembly of a tubulin oligomer in solution and this whole complex could subsequently attach to the end of the growing MT (Fig. 4b, top). Because the overall growth velocity of MTs is enhanced by the addition of even a small amount of XMAP215, this would have to mean that the XMAP215–tubulin complex has a higher affinity for the MT end than tubulin alone (possibly because of ‘pre-straightening’ of the tubulin oligomer by the XMAP215 molecule). To us, this seems a likely mechanism because it is known from electron microscopy studies that XMAP215 can bind free tubulin dimers in a protofilament-like fashion, resulting in XMAP215–tubulin complexes with lengths up to 60 nm (ref. 22). In addition, there is indirect evidence that other MT-binding proteins such as CLIP170 bind tubulin dimers before attaching to growing MTs as well^{26–28}.

The second possibility is that XMAP215 could first bind to the MT end and then accelerate the build-up of an oligomer along the length of the XMAP215 molecule (Fig. 4b, bottom). In both cases it is possible that the XMAP215–tubulin complex first binds under an angle to the MT end and then straightens out the longitudinal bond in some kind of power stroke. In fact, the instant straight addition of an oligomer 40–60 nm long through a purely brownian ratchet mechanism would require an unusually large fluctuation of the MT end away from the barrier²⁰. Finally, the step-like nature of shrinkage events indicates that XMAP215 might stay attached to the MT lattice for at least some time after arrival.

The method described here for detecting the molecular details of MT growth also paves the way for understanding the action of other classes of MT-associated proteins^{7–9}. MT-associated proteins and end-binding proteins mediate the interaction of MTs with cellular targets such as the kinetochore and the cell cortex. Understanding, at a molecular level, the operating principles of these proteins will be essential for understanding the regulation of MT dynamics in cells.

METHODS

We used an Nd:YVO₄ 1,064-nm laser (Spectra Physics) to trap a ‘construct’: a bead connected to a rigid axoneme (Fig. 1b). The laser was time-shared to create a ‘keyhole’ trap¹³, consisting of a point trap holding the bead and a line trap directing the axoneme towards a microfabricated barrier. Tests and calibrations of this set-up are given in Supplementary Methods. Photoresist (SU-8) barriers²⁹ were made by using standard microlithography techniques. Purified tubulin was obtained from Cytoskeleton Inc. and experiments were performed at 25 °C (details are provided in Supplementary Methods). Recombinant XMAP215 was expressed in Sf⁺ insect cells, purified as described previously¹⁷ and stored in liquid nitrogen. The protein concentration was determined with a Bradford assay and the molar extinction coefficient at 280 nm. Axonemes from sea urchin sperm were prepared by M. Footer from published protocols³⁰, and were bound to streptavidin-coated beads (2 µm diameter; Spherotech) by non-specific binding. Differential interference contrast (DIC) images of our experiments were recorded on DVD with a charge-coupled device camera (Kappa) and a standard DVD recorder (Philips DVD-R80). The displacement of the bead in the

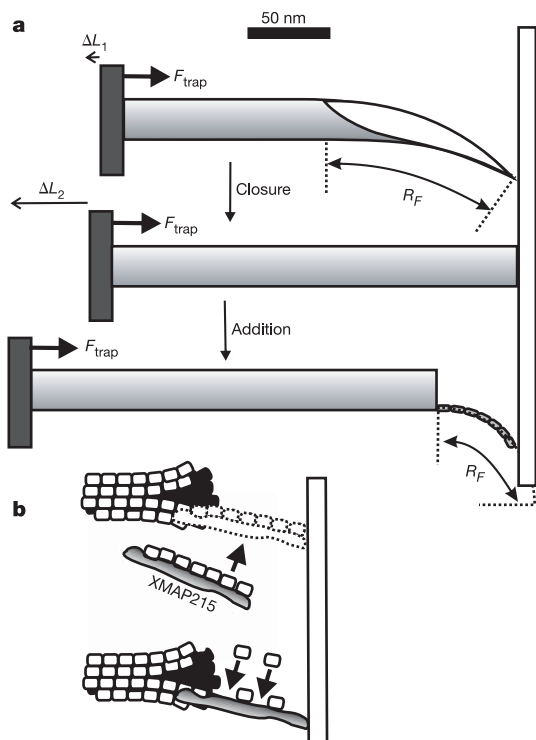


Figure 4 | MT end mechanics. **a**, Schematic drawing of a MT end under a compressive force (roughly to scale). Top: a MT with a sheet-like extension of 125 nm (preferred radius of curvature $R_0 \approx 250$ nm) adopts a force-induced radius of curvature $R_F \approx 200$ nm under a compressive force of 2.5 pN (conservative estimate of the sheet stiffness 1.2×10^5 pN nm²; that is, about 1/100 that of a MT). Middle: the same MT appears about 8 nm longer (ΔL_1) on full sheet closure. Bottom: the addition of a single, 60-nm-long protofilament (preferred radius of curvature 76 nm; stiffness 2.4×10^4 pN nm²) under a similar compressive force will lead to an apparent length increase (ΔL_2) of about 50 nm ($R_F \approx 60$ nm). **b**, Possible mechanisms for XMAP215-enhanced addition of long oligomers.

trap was measured from the digital images at a sampling rate of 25 Hz with the use of a standard auto-correlation method (image processing software home-written in IDL). A home-developed step-fitting algorithm was written in MatLab (see Supplementary Methods).

Received 3 February; accepted 5 May 2006.

Published online 25 June 2006.

- Desai, A. & Mitchison, T. J. Microtubule polymerization dynamics. *Annu. Rev. Cell Dev. Biol.* **13**, 83–117 (1997).
- Walker, R. A. *et al.* Dynamic instability of individual microtubules analyzed by video light-microscopy—rate constants and transition frequencies. *J. Cell Biol.* **107**, 1437–1448 (1988).
- Fygenson, D. K., Braun, E. & Libchaber, A. Phase diagram of microtubules. *Phys. Rev. E* **50**, 1579–1588 (1994).
- Komarova, Y. A., Vorobjev, I. A. & Borisy, G. G. Life cycle of MTs: persistent growth in the cell interior, asymmetric transition frequencies and effects of the cell boundary. *J. Cell Sci.* **115**, 3527–3539 (2002).
- Piehl, M. & Cassimeris, L. Organization and dynamics of growing microtubule plus ends during early mitosis. *Mol. Biol. Cell* **14**, 916–925 (2003).
- Kinoshita, K., Habermann, B. & Hyman, A. A. XMAP215: a key component of the dynamic microtubule cytoskeleton. *Trends Cell Biol.* **12**, 267–273 (2002).
- Schuyler, S. C. & Pellman, D. Microtubule 'plus-end-tracking proteins': the end is just the beginning. *Cell* **105**, 421–424 (2001).
- Howard, J. & Hyman, A. A. Dynamics and mechanics of the microtubule plus end. *Nature* **422**, 753–758 (2003).
- Akhmanova, A. & Hoogenraad, C. C. Microtubule plus-end-tracking proteins: mechanisms and functions. *Curr. Opin. Cell Biol.* **17**, 47–54 (2005).
- Chretien, D., Fuller, S. D. & Karsenti, E. Structure of growing microtubule ends—2-dimensional sheets close into tubes at variable rates. *J. Cell Biol.* **129**, 1311–1328 (1995).
- Wang, H.-W. & Nogales, E. Nucleotide-dependent bending flexibility of tubulin regulates microtubule assembly. *Nature* **435**, 911–915 (2005).
- Wang, H.-W., Long, S., Finley, K. R. & Nogales, E. Assembly of GMPCPP-bound tubulin into helical ribbons and tubes and effect of colchicine. *Cell Cycle* **4**, 1157–1160 (2005).
- Kersemakers, J. W. J., Janson, M. E., Van der Horst, A. & Dogterom, M. Optical trap setup for measuring microtubule pushing forces. *Appl. Phys. Lett.* **83**, 4441–4443 (2003).
- Gard, D., Becker, B. & Romney, S. MAPping the eukaryotic tree of life: Structure, function, and evolution of the MAP215/Dis1 family of microtubule-associated proteins. *Int. Rev. Cytol.* **239**, 179–272 (2004).
- Gard, D. L. & Kirschner, M. W. A microtubule-associated protein from *Xenopus* eggs that specifically promotes assembly at the plus-end. *J. Cell Biol.* **105**, 2203–2215 (1987).
- Vasquez, R. J., Gard, D. L. & Lynne, C. XMAP from *Xenopus* eggs promotes rapid plus end assembly of microtubules and rapid microtubule polymer turnover. *J. Cell Biol.* **127**, 985–993 (1994).
- Kinoshita, K., Arnal, I., Desai, A., Drechsel, D. N. & Hyman, A. A. Reconstitution of physiological microtubule dynamics using purified components. *Science* **294**, 1340–1343 (2001).
- Dogterom, M. & Yurke, B. Measurement of the force-velocity relation for growing microtubules. *Science* **278**, 856–860 (1997).
- Janson, M. E., de Dood, M. E. & Dogterom, M. Dynamic instability of microtubules is regulated by force. *J. Cell Biol.* **161**, 1029–1034 (2003).
- Janson, M. E. & Dogterom, M. Scaling of microtubule force-velocity curves obtained at different tubulin concentrations. *Phys. Rev. Lett.* **92**, 248101 (2004).
- Shirasu-Hiza, M., Coughlin, P. & Mitchison, T. Identification of XMAP215 as a microtubule-destabilizing factor in *Xenopus* egg extract by biochemical purification. *J. Cell Biol.* **161**, 349–358 (2003).
- Cassimeris, L., Gard, D., Tran, P. T. & Erickson, H. P. XMAP215 is a long thin molecule that does not increase microtubule stiffness. *J. Cell Sci.* **114**, 3025–3033 (2001).
- Janosi, I. M., Chretien, D. & Flyvberg, H. Modeling elastic properties of microtubule tips and walls. *Eur. Biophys. J.* **27**, 501–513 (1998).
- Spittle, C., Charrasse, S., Larroque, C. & Cassimeris, L. The interaction of TOGP with microtubules and tubulin. *J. Biol. Chem.* **275**, 20748–20753 (2000).
- VanBuren, V., Cassimeris, L. & Odde, D. J. A mechanochemical model of microtubule structure and self-assembly kinetics. *Biophys. J.* **89**, 2911–2926 (2005).
- Diamantopoulos, G. S. *et al.* Dynamic localization of CLIP-170 to microtubule plus ends is coupled to microtubule assembly. *J. Cell Biol.* **144**, 99–112 (1999).
- Arnal, I., Heichette, C., Diamantopoulos, G. S. & Chretien, D. CLIP-170/tubulin-curved oligomers coassemble at microtubule ends and promote rescues. *Curr. Biol.* **14**, 2086–2095 (2004).
- Folker, E. S., Baker, B. M. & Goodson, H. V. Interactions between CLIP-170, tubulin, and microtubules: implications for the mechanism of CLIP-170 plus-end tracking behaviour. *Mol. Biol. Cell* **16**, 5373–5384 (2005).
- Schek, H. T. III & Hunt, A. J. Micropatterned structures for studying the mechanics of biological polymers. *Biomed. Microdevices* **7**, 41–46 (2005).
- Pierce, D. W. & Vale, R. D. Assaying processive movement of kinesin by fluorescence microscopy. *Methods. Enzymol.* **298**, 154–171 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Hyman and T. Mitchison for discussions; K. Kinoshita for help with the purification of XMAP215; S. Tans, K. Kuipers and D. Drechsel for a critical reading of the manuscript; and M. Footer for the gift of axonemes. This work is part of the research program of the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is supported financially by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.D. (dogterom@amolf.nl).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

For scientists, political, social and religious differences tend to fall by the wayside in the pursuit of data. Scientific collaborations usually transcend such boundaries. A letter I received recently reminded me of these ideals — and emphasized how sorely they are being tested in the Middle East at the moment.

The letter was from Zachary Lippman, a biology postdoc based at the Hebrew University of Jerusalem. When he accepted a fellowship to work in the laboratory of Dani Zamir, Lippman wasn't expecting politics to have any impact on his research. But it has. The plants for Lippman's genetics experiments are in fields in Acre, a few kilometres south of the Israeli-Lebanese border. Since the conflict between Israel and Hezbollah erupted, missiles have fallen within 500 metres of the field where Lippman and his labmates collect data with Arab helpers from a nearby village.

When the fighting first broke out, Lippman's first thoughts were about how the conflict might affect his research. Now, he is asking larger questions about science, politics and war, as members of his lab are being called up into Israel's army.

"It is startling that my graduate-student colleagues are forced to exchange their lab books for weapons with a single phone call, and that the very land on which my plants are growing is the source of suffering for both sides of the conflict," Lippman writes. As more reservists are called up, Lippman thinks of the graduate student who sat next to him, and who is currently in Lebanon. "A graduate student's first commitment is not, necessarily to his research," Lippman notes. "Rather, it is to the military, and this is often not by choice."

Lippman has vowed to stick with his research, even if it means repeating some of his experiments. He also aims to maintain collaboration with both Arabs and Israelis. "I hope to see my colleagues back in the lab soon, fighting the logical scientific battles in which they were previously immersed, instead of the illogical type they are fighting now," he writes.

Paul Smaglik, *Naturejobs* editor

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestranguez (4994)

UK/Ireland/Italy/RoW:
Loredana Milanese (4944)
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production:
Catherine Alexander

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

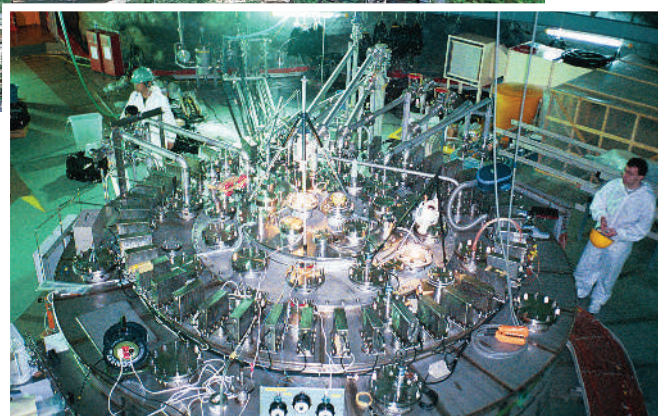
Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com



CITY OF SENDAI

Japan's other research hub

The city of Sendai has much to offer research and industry, but, says **David Cyranoski**, competition for funding and brains is stiff.



Particle physicists in Sendai (top) have led international work on the high-calibre KamLAND antineutrino detector.

In Japan, Sendai is known for its lush greenery, its nearby hot springs and its beef tongue. Outside Japan, it is hardly known at all. Sendai has long lived in the shadow of the Tokyo/Yokohama hub in the Kanto plain and the Osaka/Kyoto/Kobe triangle that anchors the Kansai region. But it has been responsible for seminal advances in materials science, microelectronics and particle physics. Its researchers, particularly those at the city's core institution, Tohoku University, have a track record for turning basic science into practical, profitable applications. And the university's involvement with world-class facilities, such as the KamLAND antineutrino detector at Kamioka, attract international attention. Meanwhile, local initiatives are helping to attract companies and bring scientific understanding to the public.

All Japanese universities are facing increasing competition and expectations to find practical applications for their basic research. For Sendai to carve a niche, Tohoku University and its neighbouring research institutions will need to work closely with the local business community, launch new projects that bring global collaboration, and overcome the obstacles that make it harder for women (see 'Women's work') and Westerners to work in Japan.

Sendai already has a solid foundation on which to build, says Richard Dasher, director of the US-Asia

Technology Management Center at Stanford University, California. The Sendai region boasts 13 universities, 5 junior colleges and 2 technical colleges. Overall, these have a total of 80,300 students. Tohoku University alone has more than 2,600 research and teaching staff, some 85% of whom work in the sciences. "If the research there was happening in Tokyo or Kyoto, it would be getting a lot more press," says Dasher.

Competition for the best brains is growing, as the number of students entering Japanese universities has dropped. A nationwide reorganization of the university system in 2004 (see *Nature* **429**, 210–214; 2004) has made it easier for faculty members to move from one institution to another, and for universities to compete for funds. The economic and industrial threat of China and other neighbours also looms, says the city's new mayor, Katsuhiko Umehara. "Global competition is becoming tougher and tougher, so Sendai will need to have a good strategy to compete with other cities, not only in Japan, but also in east Asia and the rest of the world."

However, Sendai has some advantages over its more prominent neighbours. A bounty of trees provides an alternative to the concrete and neon of other Japanese metropolises. It is only a two-hour bullet-train journey from Tokyo, but the buffer zone this distance creates provides the perfect setting for creative work, says mathematician Motoko Kotani. "Somehow, in Tohoku,

Sendai's mayor, Katsuhiko Umehara, is aware that the city faces tough competition from high-tech hubs within Japan and around the world.



LAWRENCE BERKELEY NATIONAL GALLERY

I feel time goes slowly and quietly, and people can keep some distance from current trends. You can wait patiently for your idea to ripen,” she says.

Tohoku University’s greatest competitive advantage is its tradition of successful applied research. In the 1920s, two Tohoku University professors invented the Yagi-Uda antenna. This has many applications, one of the most recognizable of which is as a roof-top television antenna. And research dating back to the 1950s led to the development in the 1970s of perpendicular magnetic recording, a technology that is now poised to revolutionize hard disk drives with its tremendous memory capacity.

Ahead of the game

Many current research projects hold great, if underdeveloped, potential for future industrial application. Researchers at the university’s best-known research organization, the Institute of Materials Research (IMR), are trying to turn expertise in collecting and analysing data on combinatorial materials into a new field called materials informatics.

Hideo Ohno, who heads the university’s Laboratory for Nanoelectronics and Spintronics, says the key to these Tohoku achievements has been a tradition of going back to the fundamentals in parallel with working closely with industry. Ohno’s own discovery — that electricity can be used to control the magnetic properties of semiconductors — led to the development of low-power-consumption semiconductor devices. “Researchers are aware of the needs of those outside academia even though they are working on basic research,” he says. Likewise, Tadahiho Ohmi’s Fluctuation-Free Facility for Information Industry lists a dozen core companies that are participating in his efforts to develop malfunction-free sub-100-nanometre semiconductors and other semiconductor technologies. “We try to stay 15 years ahead of industry, as we will be leading it into the future,” says Yasuyuki Shirai, an assistant professor at the facility.



City-wide, regional and international collaborations are trying to make the most of this cutting-edge research. In 2002, the education ministry designated Sendai an ‘Intelligence Knowledge Cluster’ as part of an initiative that gives the city about ¥500 million (US\$4.35 million) per year for five years to develop ‘intelligent electronics’. The dozen joint-research projects between industry and Japanese corporations are aimed directly at society’s welfare — for example, monitoring devices that accurately estimate energy consumption to help in the treatment of diabetic patients.

Some of these projects are expected to boost the Sendai–Finland Wellbeing Center, a collaboration established in 2005 to promote the self-reliance of one of Japan’s largest and fastest growing markets — the elderly. The 12 research projects are dominated by Tohoku University researchers, but two projects are being led by Sendai National College of Technology and one by Tohoku Institute of Technology.

Gathering steam

The trade ministry is also funding regional cluster programmes. Since 2001, it has given about ¥700 million per year to a project involving hundreds of academic organizations and small to medium-sized businesses that endeavours to turn successful academic work into commercial technologies aimed at helping Japan’s ageing population. About ¥380 million per year has been allotted to a similar consortium of academic and industrial organizations looking for technologies to promote a sustainable society.

Some supporting mechanisms are already in place. The Intelligent Cosmos Research Institute has had some success in attracting research and development companies to the region since 1989. In 2004, the ¥3.2-billion Tohoku Incubation Fund started investing in university ventures. Sendai city is also converting an 81-hectare golf course into a science park in the hope that 50 companies, from inside and outside Japan, will set up business there.

WOMEN’S WORK

Mathematician Motoko Kotani (pictured) rose to fame in Japan without facing much discrimination. But a star female researcher is still a rare event in Japan, and Kotani and others at Tohoku University have set out to level the playing field.

At present, 11.9% of Japanese



scientists are female, a figure well below that of most industrialized countries. The central government has struggled to take greater advantage of female brainpower with various policies, but so far results have been modest.

Kotani, who is involved in an initiative to help Japanese women scientists advance their careers, says the key is helping women over certain stumbling blocks. These include encouraging female students to consider a career in science, helping researchers balance the demands of childbirth and child rearing with their careers, and ensuring that women do not feel overwhelmed by their minority status.

Tohoku University has historically been a leader. In 1913, it became the first of Japan’s imperial universities

to allow female students. In 2002, it established the Sawayanagi Prize, which is awarded to a successful research collaboration between women and men.

But Tohoku University remains below the national average: only 7.7% of faculty members, and only 6.3% of science faculty members, are female. Starting this year, Tohoku established policies to change this. Nurseries have been set up to care for children, and provide medical treatment when children are sick, and part-time workloads and support staff are available to working mothers.

Most notably, Tohoku University has established the ‘science angels’, a group of 40 graduate students who visit schools and encourage students to become scientists, thereby providing much-

needed role models. The science angels also work to bring women at the university together.

The goal is to have women make inroads at all levels. By 2020, for example, they hope to have 30% of all committee seats and senior positions filled by women.

Kotani, who last year won Japan’s Saruhashi Prize, which recognizes the work of Japanese women in science, says the solitary, theoretical nature of her research might have helped her avoid the more egregious forms of discrimination that a woman on a large experimental team might experience. “In maths, you don’t need any facilities — only talent — and you are judged by what you have done alone,” she says. She hopes the Tohoku project will extend this freedom to all Japanese women scientists.

D.C.

But it remains unclear how much the local economy or Tohoku University will benefit from even the highest-impact research. Venture-capital support for local businesses lags behind that of other research hubs such as Osaka. Many scientists are happy to stick to basic research questions or collaborate with researchers far and wide — neither of which contributes to building a foundation for local industry. The semiconductor fabrication research at Ohmi's laboratory has been tremendously influential since the 1980s, but has generated little profit for the university as the group has tended to share their protocols freely with others in the field. "We want the technology to be used anywhere that researchers can take advantage of it," says Shirai.

Liquid asset

Access to some unique facilities is another of Sendai's great draws. Tohoku particle physicists, for example, have led an international collaboration of scientists from 13 institutions on the KamLAND antineutrino detector. The detector has a 1,000-tonne tank holding a liquid whose molecules emit light when antineutrinos collide with them. KamLAND is 400 kilometres from Sendai and is buried in an old mine shaft to screen out other subatomic particles. Machines as extensive and well-protected as this are rare, with only a handful worldwide, and this team has taken advantage of its location to discover new aspects of antineutrinos emitted from Japan's various nuclear reactors. The team has developed the field of neutrino geophysics, which uses the detection of antineutrinos released by radioactive elements within Earth to answer questions about Earth's heat balance.

The microelectromechanical systems (MEMS) facilities headed by Masayoshi Esashi also draw international and domestic researchers. But many of the 350 researchers — including 20 foreigners — who have come to Sendai from 39 labs are hoping for profit beyond publication. Esashi says that 35 years of experience is being brought to bear on MEMS, which integrate sensors and other electrical equipment on a single chip. "The knowledge is managed to be accessible, which I think should be the mission of the university," he says.

These ties give good reason for Sendai to be optimistic about possible industrial applications.

"In a university system increasingly being modelled on the US system, some natives still find the Japanese way a boon to research."

AT HOME IN SENDAI

Japan has a reputation for being a difficult place for non-Japanese scientists to work, in terms of language, culture and the different ways in which research is organized. But some outsiders have adapted to Sendai.

Richard Smith, a supercritical-fluids specialist at Tohoku University, has taught physics classes in English with some success. "The students study hard — if you give them the materials beforehand, they will come prepared," he says.

Daily life necessities, such as English-speaking doctors, are also accessible, says Alexandre Kozlov, a Russian-born Australian neutrino researcher whose

second child was born in Japan. Kozlov says it was a joy to work in the orderly and efficient Japanese group on the KamLAND neutrino projects. "Each member could have done various parts of the experiment, but they each knew their role and did it well," he says.

Rather than being a barrier, some foreign researchers say the cultural aspects are a big draw. Smith says the *koza* or chair system, in which a strict hierarchy is maintained in the laboratory, is an essential foundation for mutual support and mentoring among his junior faculty and students. "They wouldn't know what to do without it," he says.

D.C.



The tranquil surroundings of Tohoku University offer a perfect setting for creative thought.

TOHOKU UNIVERSITY

Esashi has spun off a company called MEMS Core, and a 'MEMS Park Consortium', which was established in 2004 with Sendai's assistance, has gathered more than 110 members.

And in a university system increasingly being modelled on the US system, some natives still find the Japanese way a boon to research. "People do not move from company to company in Japan, so companies do not mind dispatching them to the university. They will come back," says Esashi. "We are taking advantage of Japanese work habits to get a good collaboration."

Researchers are also buoyed by positive support from the community. Being recognized and accepted as a scientist makes adapting to life in Japan easier, says Alexandre Kozlov, a Russian-born Australian neutrino researcher (see 'At home in Sendai').

The city of Sendai has worked with the university to cultivate this awareness. A monthly science café is held in the architecturally innovative Sendai Mediatheque, a seven-storey structure with irregular tubes running through it that house elevators and staircases. The monthly events introduce researchers' work to the hundred or so high-school students and other local residents that show up. Kotani's lecture was introduced by Sendai's mayor, who discussed the significance of mathematics using Fermat's theorem as an example.

Dasher is concerned that, despite support from the city and Sendai's various attractions, many students leave the Tohoku region for industry jobs in Tokyo or Osaka after a BS or MS degree. He thinks that changes in Japan's economy will enhance the value of PhD holders and university research in general — drawing more attention from students and industry alike to Tohoku University's fields of expertise.

If predictions prove right, Sendai and Tohoku University will shine. Dasher, who now works as a special adviser to the president of Tohoku University, says: "It reminds me of Silicon Valley before it became known as Silicon Valley."

David Cyranoski is Nature's Asia-Pacific correspondent.

Statler pulchrifex

Make love *and* war.

Matt Weber

"Did I ever tell you," asked Statler, "about the time we created the most beautiful woman in the world?"

"That was us?" said Waldorf. He bowled a blue rock, which knocked Statler's red one off the board and came to rest in the three-point zone. "I thought it was MI6 for sure."

Waldorf, short and bald, had coordinated covert missions with the Army Rangers and several more obscure units; Statler, rangy and jowly, was secretive about his affiliation. They were shuffleboard buddies and imaginative hecklers of the home's cafeteria food. In the wake of a particularly inexcusable meatloaf night, Ethel R. had named them Statler and Waldorf after the Muppet hecklers; everyone had code names anyway, so the nicknames had stuck. It was the end of a warm afternoon.

"You and everyone else," said Statler. "Come on, some Oxbridge teabag's gonna engineer the most beautiful woman in the world? You need lust for that. You need to bathe in meat and oil and women. When you've spit off the edge of what human experience has to offer, then you're ready to build the most beautiful woman in the world. Anything less, you'll go running home to mama at the first teratogenesis." Statler bowled his blue rock, which sailed past the end of the board to join the one Waldorf had knocked out. "Didn't you ever take no psychology?"

"I never learned that in class, anyway. D'you weaponize her?"

"I look like a goofball to you? Look, we spent millions on testing alone. You gotta find the subjects, you gotta get the electrodes in them, you gotta keep their paws off her, you gotta get enough data to swear up and down that the sight, the mere sight, of this piece is driving these idiots' limbic systems to the edge of the refractory period. Of course we weaponized her."

Waldorf sent a red rock into the two-point zone. "I thought the problem with weaponizing beauty was the countermeasure. Just whip up some eunuch type and dial down the gain in the limbic units. Bang — no aesthetic response."

"Dial down the gain," snorted Statler. "Look, you could depress the synaptic strengths so your eunuch wouldn't step

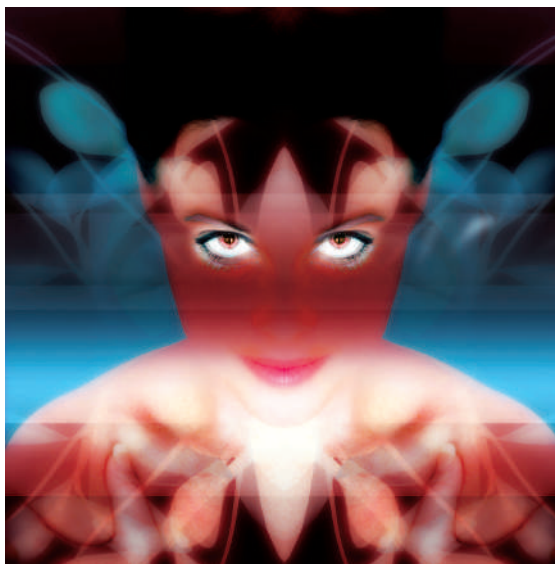
aside for a charging bull elephant and Miss Mata would still send their firing rates to the wall. I repeat, I look like a goofball to you?"

"You do, since you're so interested," said Waldorf. "So what next?"

"The usual song and dance. Cloning, ninja training, brainwashing. You probably had some guys on it."

Statler bowled a blue rock at Waldorf's guard rock and missed. Waldorf snorted a laugh.

"Back then I could have filled out three forms to have you killed," said Statler. "For your immediate family, two. I know for a



fact you'd have needed six to snuff me, and I'd have had my eye on four of them. You'd have disappeared before they cleared the comptroller."

Statler hated losing at shuffleboard.

"Well," said Waldorf, "we're not sending clones of Spooksville's Next Top Model to do all our black ops. So what went wrong?"

Statler scowled. "Nothing at first. But some genius in Delhi got wind of the project. Figured, what with the way they were snarling at Pakistan, we might send Mata to do some sabotage on their weapons systems, keep the temperature down. He was right, of course. And since he" — Statler missed Waldorf's guard rock miserably — "was smarter than you, when he built the countermeasure, he did not dial down the gain. He dialled it up."

"Dialled it up." Waldorf polished the sweat from his pate. "The eunuch loved everything?"

"He had his balls, but yeah. To him, everything was as beautiful as Mata. He could barely eat for weeping and laughing. They called him the Bodhisattva Assassin."

Waldorf's rock went short, alighting just shy of the two-point line. "You're telling me some chortling Hare Krishna beat a Ranger-trained, brainwashed, sex-sweating beauty machine?"

"Sure," said Statler. "He loved watching things die exactly as much as he loved anything else. That was the point of him."

Waldorf nodded.

"Imagine you go through life," Statler

added, "knowing any man or woman will at least gasp and stare before trying to kill you. Every trainer, every superior you've had reinforces this impression. You get to expect breathing room. That's all it takes." Statler shook his head. "I saw the tape. Vicious stuff. He was sobbing for joy the whole time." His next bowl knocked both of Waldorf's three-point rocks from the board. "We sent a few more of her after him, just to see. Project Mata Hari, zero; transgenic Buddha, eight. So we decommissioned her, and the Bodhisattva Assassin died of a heart attack at 19, and India nuked Pakistan. The End."

"Well," said Waldorf, "it worked out all right." He squinted up at Statler. "You know a lot about Delhi's side of the story."

Statler tilted his head towards an old man standing very still on the tenth hole of the golf course. "Sanskrit was the genius behind the Bodhisattva Assassin. Told me all about it over bridge."

"How'd he get in here, then?"

"Defected."

"Huh," said Waldorf. "I always thought Sanskrit had something under the hood."

"Smarter than you, anyway." Statler snorted. "Dial down the gain." His final shot knocked his two-point rock to the very edge of the three-point zone. "I win."

The clear sky was darkening, the warm air gently cooling. Waldorf and Statler collected their rocks and deposited them in the wicker basket by the shuffleboard. They entered the cafeteria with the peaceful anticipation of men whose wars had already been fought.

Matt Weber is a graduate student in cognitive neuroscience at Princeton University, New Jersey.

JACEY